

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 01:50:51 ; Search time 3558.48 seconds  
(without alignments)  
11163.571 Million cell updates/sec

Title: US-09-986-682B-4

Perfect score: 1365  
Sequence: 1 ATGAACAGCGGAGCTACCA.....CCGACAGAGCAATTCCTCA 1365

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pal:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_on:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1365	100.0	1365	6	AR078181	AR078181 Sequence
2	1365	100.0	1365	6	AR208487	AR208487 Sequence
3	1365	100.0	1365	6	E15382	E15382 Bacillus sp
4	1365	100.0	2408	1	AB010272	AB010272 Bacillus
5	1365	100.0	2408	6	AR078182	AR078182 Sequence
6	1365	100.0	2408	6	AR208488	AR208488 Sequence
7	1365	100.0	2408	6	E15383	E15383 Bacillus sp
8	735.8	55.4	11318	1	AE007686	AE007686 Clostridi
9	617.2	45.2	1449	6	AX434229	AX434229 Sequence
10	578.6	42.4	2007	1	BSSACB6	X02730 Bacillus su
11	578.6	42.4	5831	12	AF047518	AF047518 Cloning v
12	578.6	42.4	5842	12	AF004910	AF004910 Cloning v
13	578.6	42.4	5846	12	XXU17500	U17500 Cloning vec
14	578.6	42.4	5941	12	SYNPRC	L05083 Cloning vec
15	578.6	42.4	6347	12	SYNPRC	L05081 Cloning vec
16	578.6	42.4	6349	12	AF047519	AF047519 Cloning v
17	578.6	42.4	6798	12	SYNPRB	L05082 Cloning vec
18	578.6	42.4	9297	12	AF048702	AF048702 Site-spec
19	578.6	42.4	88063	1	BS294043	Z94043 B. subtilis
20	578.6	42.4	209510	1	BS080018	Z99121 Bacillus su
21	577	42.3	10317	6	AR162197	AR162197 Sequence
22	565.8	41.5	1422	1	BSU34874	U34874 Bacillus st
23	542.8	39.8	2350	1	BASACB	X52988 Bacillus am
24	542.8	39.8	11612	12	CVU80929	U80929 Cloning vec
25	542.8	39.8	16011	12	XXU09128	U09128 psacB11 pl
26	542.8	39.8	16297	12	CVU75991	U75991 Cloning vec
27	542.8	39.8	18772	12	AF133437	AF133437 Cloning v
28	542.8	39.8	19500	12	CVU75992	U75992 Cloning vec
29	542.8	39.8	22530	12	AB020028	AB020028 Transform
30	542.8	39.8	107872	2	AC013485	AC013485 Trypanoso
31	542.8	39.8	110000	2	AC058788_2	Continuation (3 of
32	542.8	39.8	113316	2	AC007866	AC007866 Trypanoso
33	542.8	39.8	122526	2	AC012647	AC012647 Trypanoso
34	542.8	39.8	132928	2	AC091089	AC091089 Oryza sat
35	542.8	39.8	133901	2	AC008146	AC008146 Trypanoso
36	542.8	39.8	146349	2	AC007863	AC007863 Trypanoso
37	542.8	39.8	151443	2	AC007865	AC007865 Trypanoso
38	542.8	39.8	154413	2	AC007862	AC007862 Trypanoso
39	542.8	39.8	155204	2	AC007926	AC007926 Trypanoso
40	542.8	39.8	156128	2	AC113265	AC113265 Mus muscu
41	542.8	39.8	156709	2	AC013484	AC013484 Trypanoso
42	542.8	39.8	156754	2	AC008031	AC008031 Trypanoso
43	542.8	39.8	157774	2	AC013353	AC013353 Trypanoso
44	542.8	39.8	159717	2	AC008368	AC008368 Trypanoso
45	542.8	39.8	162249	2	AC009259	AC009259 Trypanoso

#### ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	FEATURES
AR078181	AR078181	Sequence 4 from patent US 5962297.	AR078181	AR078181.1	GI:10004927	Unknown.	Unknown.	1 (bases 1 to 1365)	Tsukaki,K., Kubota,M. and Chaen,H.	Polypeptides having .beta.-fructofuranosidase activity	Patent: US 5962297-A 4 05-OCT-1999;
											Location/Qualifiers

Prod. No. is the number of results predicted by chance to have a

source 1. 1365  
/organism="unknown"  
BASE COUNT 459 a 266 c 283 g 357 t  
ORIGIN

Query Match 100.0%; Score 1365; DB 6; Length 1365;  
Best Local Similarity 100.0%; Pred. No. 1.7e-305;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGAACAGCGGGGACTACAAAGAGACGTATGTTTTGCCATATTAACAGCGCTGACATG 60
DB 1 ATGAACAGCGGGGACTACAAAGAGACGTATGTTTTGCCATATTAACAGCGCTGACATG 60
OY 61 CTAATAATTCAGAGCAACAAAACAGTCCCTCAATTTAAAGTCCCTCAATTCATGATCA 120
DB 61 CTAATAATTCAGAGCAACAAAACAGTCCCTCAATTTAAAGTCCCTCAATTCATGATCA 120
OY 121 GCAATCAAAAACATTGATTCGGCAAAAGGCTATGATAGTCAGGCACTTAATAGATTTA 180
DB 121 GCAATCAAAAACATTGATTCGGCAAAAGGCTATGATAGTCAGGCACTTAATAGATTTA 180
OY 181 GATGATGGGATAGTGGCCACTGCAAAAACGCTGATGCTAGCGGCAATTAATGATGA 240
DB 181 GATGATGGGATAGTGGCCACTGCAAAAACGCTGATGCTAGCGGCAATTAATGATGA 240
OY 241 TATCAGATCGTCCGCTTTAGCAGGTGACCCAAAAACAGTATGATCTCCACTTCAT 300
DB 241 TATCAGATCGTCCGCTTTAGCAGGTGACCCAAAAACAGTATGATCTCCACTTCAT 300
OY 301 TTATCTCATCAAAAAGTCGGTATACATGATTTGACAGCTGCAAAAATCTGGAAGATTA 360
DB 301 TTATCTCATCAAAAAGTCGGTATACATGATTTGACAGCTGCAAAAATCTGGAAGATTA 360
OY 361 TTTGAAGATATGATAAATTTGTTCCAAATGATCCGTATCTTAATATCAACACAGGAG 420
DB 361 TTTGAAGATATGATAAATTTGTTCCAAATGATCCGTATCTTAATATCAACACAGGAG 420
OY 421 TGGTCAGGTTCTGCTACTTTAAACAAAGATGCGCAAGTCCGTTTATCTATACAGATTAC 480
DB 421 TGGTCAGGTTCTGCTACTTTAAACAAAGATGCGCAAGTCCGTTTATCTATACAGATTAC 480
OY 481 TCAGTAAATCCCTGAAGATGCTGGAACCGGTCTGTTGTAACCAATCTTTCAACTGCTCA 540
DB 481 TCAGTAAATCCCTGAAGATGCTGGAACCGGTCTGTTGTAACCAATCTTTCAACTGCTCA 540
OY 541 GTAACCTTATCCAGCGGATGACAGTACCTTAAGTGAATGATGATCATATAA 600
DB 541 GTAACCTTATCCAGCGGATGACAGTACCTTAAGTGAATGATGATCATATAA 600
OY 601 TCTGCTTTGATGCGGAGACGCTACAGTTTATCAAAATATTCAGCAATTTATCGATGA 660
DB 601 TCTGCTTTGATGCGGAGACGCTACAGTTTATCAAAATATTCAGCAATTTATCGATGA 660
OY 661 GCGAAGTGGATTTCAGGTATTAACCATCTTTAAAGGACCTCAGTATGTTAAAGATAAG 720
DB 661 GCGAAGTGGATTTCAGGTATTAACCATCTTTAAAGGACCTCAGTATGTTAAAGATAAG 720
OY 721 GGCCTAATAATATCTGTCTTTGAAGGAATTAAGTGAACAGATGATGATGAAGCGCAT 780
DB 721 GGCCTAATAATATCTGTCTTTGAAGGAATTAAGTGAACAGATGATGATGAAGCGCAT 780
OY 781 CAGTCTTCAATTAATTAAGCTTACTATGCGGAGAGTCTTCTTCAGAGATGAATAA 840
DB 781 CAGTCTTCAATTAATTAAGCTTACTATGCGGAGAGTCTTCTTCAGAGATGAATAA 840
OY 841 AATAACAGCTTCAAGTCTTAATAAACAATTCCTTTTAGCGAATGCTCATTTAGGC 900
DB 841 AATAACAGCTTCAAGTCTTAATAAACAATTCCTTTTAGCGAATGCTCATTTAGGC 900
OY 901 ATTGTTGAATGGCGGATGACTATACAGTGAAGAGTGTATGAACCATTAATGATCATCA 960
DB 901 ATTGTTGAATGGCGGATGACTATACAGTGAAGAGTGTATGAACCATTAATGATCATCA 960

```

OY 961 AACACAGTAGCAGATGAAGTGCAGACGCGCAATATATTTAAATGAATTAATGATAT 1020  
DB 961 AACACAGTAGCAGATGAAGTGCAGACGCGCAATATATTTAAATGAATTAATGATAT 1020  
OY 1021 CTATTCACGGATTTCAGAGATTCCAAAATGACAGTATGATGATTTAACGACAAAGATGTT 1080  
DB 1021 CTATTCACGGATTTCAGAGATTCCAAAATGACAGTATGATGATTTAACGACAAAGATGTT 1080  
OY 1081 TATATGCTAGAGCGCGGAGCGGACATCTTAATATGAGCCCAACACCGGTAATGAAT 1140  
DB 1081 TATATGCTAGAGCGCGGAGCGGACATCTTAATATGAGCCCAACACCGGTAATGAAT 1140  
OY 1141 GCACCTGTAATTAACATGAATCTTGACCCCTGCTGATCTCACACACACTTACTCTATTGC 1200  
DB 1141 GCACCTGTAATTAACATGAATCTTGACCCCTGCTGATCTCACACACACTTACTCTATTGC 1200  
OY 1201 GGTATCCCGCACCCCTGGAAGTATATATGTTGTTCTACACAAAGTTATATGACGAATGAGGC 1260  
DB 1201 GGTATCCCGCACCCCTGGAAGTATATATGTTGTTCTACACAAAGTTATATGACGAATGAGGC 1260  
OY 1261 TTCTATCCAGACATCACTCTCACCCTGCGGAGCAAGCTTGGGGTTAATATTAAGGCTCT 1320  
DB 1261 TTCTATCCAGACATCACTCTCACCCTGCGGAGCAAGCTTGGGGTTAATATTAAGGCTCT 1320  
OY 1321 GACACATCTGAGAGAAAATATGTTCCGACAAAGACAATTCOCA 1365  
DB 1321 GACACATCTGAGAGAAAATATGTTCCGACAAAGACAATTCOCA 1365

RESULT 2  
AR208487 1365 bp DNA linear PAT 20-JUN-2002  
LOCUS AR208487  
DEFINITION Sequence 4 from patent US 6383769.  
ACCESSION AR208487  
VERSION AR208487.1 GI:21509653  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1365)  
AUTHORS Tsusaki, K., Kubota, M. and Chaen, H.  
TITLE Polypeptides having .beta.-fructofuranosidase activity  
JOURNAL Patent: US 6383769-A 4 07-MAY-2002;  
FEATURES  
source 1. 1365  
/organism="unknown"  
BASE COUNT 459 a 266 c 283 g 357 t  
ORIGIN

Query Match 100.0%; Score 1365; DB 6; Length 1365;  
Best Local Similarity 100.0%; Pred. No. 1.7e-305;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGAACAGCGGGGACTACAAAGAGACGTATGTTTTGCCATATTAACAGCGCTGACATG 60
DB 1 ATGAACAGCGGGGACTACAAAGAGACGTATGTTTTGCCATATTAACAGCGCTGACATG 60
OY 61 CTAATAATTCAGAGCAACAAAACAGTCCCTCAATTTAAAGTCCCTCAATTCATGATCA 120
DB 61 CTAATAATTCAGAGCAACAAAACAGTCCCTCAATTTAAAGTCCCTCAATTCATGATCA 120
OY 121 GCAATCAAAAACATTGATTCGGCAAAAGGCTATGATAGTCAGGCACTTAATAGATTTA 180
DB 121 GCAATCAAAAACATTGATTCGGCAAAAGGCTATGATAGTCAGGCACTTAATAGATTTA 180
OY 181 GATGATGGGATAGTGGCCACTGCAAAAACGCTGATGCTAGCGGCAATTAATGATGA 240
DB 181 GATGATGGGATAGTGGCCACTGCAAAAACGCTGATGCTAGCGGCAATTAATGATGA 240
OY 241 TATCAGATCGTCCGCTTTAGCAGGTGACCCAAAAACAGTATGATCTCCACTTCAT 300
DB 241 TATCAGATCGTCCGCTTTAGCAGGTGACCCAAAAACAGTATGATCTCCACTTCAT 300

```

```

OY 301 TTATCTATCAAAAAGCGGTGATACATGATTCAGAGCTGGAAAAATGCTGGAGAGTA 360
DB 301 TTATCTATCAAAAAGCGGTGATACATGATTCAGAGCTGGAAAAATGCTGGAGAGTA 360
OY 361 TTTGAAGATATGATTAATTTGTTCCTAATGATCCGTATCTTAATATCAACACAGAG 420
DB 361 TTTGAAGATATGATTAATTTGTTCCTAATGATCCGTATCTTAATATCAACACAGAG 420
OY 421 TGTGAGGTCTGCTACTTAAACCAAGATGCGGAGCTGCTTTCATGATCAGATTAC 480
DB 421 TGTGAGGTCTGCTACTTAAACCAAGATGCGGAGCTGCTTTCATGATCAGATTAC 480
OY 481 TCAGGTAATCTCGAAGATGCTGGAACCGGTGCTGTAACCAATATCTTCACTGCTCA 540
DB 481 TCAGGTAATCTCGAAGATGCTGGAACCGGTGCTGTAACCAATATCTTCACTGCTCA 540
OY 541 GTAACCTATATCCAGCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 541 GTAACCTATATCCAGCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
OY 601 TCTGCTTGTGATGCGGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 601 TCTGCTTGTGATGCGGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 660
OY 661 GCGCAAGTGGATTTGAGTATACCATCTTAAAGACCCCTACTATGTTGAAGATAG 720
DB 661 GCGCAAGTGGATTTGAGTATACCATCTTAAAGACCCCTACTATGTTGAAGATAG 720
OY 721 GGGCAATATATCTTGTCTTTGAGAGCAATCTGGAACACAGATGTTATCAAGCCAT 780
DB 721 GGGCAATATATCTTGTCTTTGAGAGCAATCTGGAACACAGATGTTATCAAGCCAT 780
OY 781 CAGTCTTCAATATTAAGCTTACTATGCGGAGAGCTGCTTCTTCAGAGATGAAAA 840
DB 781 CAGTCTTCAATATTAAGCTTACTATGCGGAGAGCTGCTTCTTCAGAGATGAAAA 840
OY 841 AATAAAGCTCTTCAAGTCTTAAAAACCAATTTGCTTTAGCGAATGTCATTAGGC 900
DB 841 AATAAAGCTCTTCAAGTCTTAAAAACCAATTTGCTTTAGCGAATGTCATTAGGC 900
OY 901 ATTGTTGAATGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 901 ATTGTTGAATGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
OY 961 AACACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 961 AACACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
OY 1021 CTATTCACGATTCGAAGAGATCCAAATGACAGATGATGATGATGATGATGATGAT 1080
DB 1021 CTATTCACGATTCGAAGAGATCCAAATGACAGATGATGATGATGATGATGATGAT 1080
OY 1081 TATATGCTAGGCGCGGAGGCGATCTTAAATGCGCCACACACACCCGATTAATGAACT 1140
DB 1081 TATATGCTAGGCGCGGAGGCGATCTTAAATGCGCCACACACACCCGATTAATGAACT 1140
OY 1141 GCACTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 GCACTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
OY 1201 GGTATCCGCAACCTGAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 GGTATCCGCAACCTGAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1260
OY 1261 TTTCTATCCAGAAATCACTCTACCTGCGGAGCAAGCTTGGGTTAATTAAGGCTCT 1320
DB 1261 TTTCTATCCAGAAATCACTCTACCTGCGGAGCAAGCTTGGGTTAATTAAGGCTCT 1320
OY 1321 GACACATCTGGAGAGAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1365
DB 1321 GACACATCTGGAGAGAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1365

```

```

RESULT 3
E15382
LOCUS Bacillus sp. gene for beta-fructofuranosidase. 1365 bp DNA linear PAT 28-JUL-1999
DEFINITION E15382
ACCESSION E15382
VERSION E15382.1 GI:5710065
KEYWORDS JP 199806586-A/1.
SOURCE Bacillus sp.
ORGANISM Bacillus sp.
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 1365)
AUTHORS Tsusaki, K., Kubota, M., and Chaen, H.
TITLE POLYPEPTIDE HAVING BETA-FRUCTOFURANOSIDASE ACTIVITY
JOURNAL Patent: JP 199806586-A 1 10-MAR-1998.
HAYASHIBARA BIOCHEM LAB INC
COMMENT OS Bacillus sp.
PN JP 199806586-A/1
PD 10-MAR-1998
PE 09-JUN-1997 JP 1997164875
PR 10-JUN-1996 JP 96P 170630
PI TSUSAKI KEIJI, KUBOTA MICHIO, CHAEN HIROKO
PC C12N15/09, C07H21/04, C07K14/32, C12N1/21, C12N9/10, C12N9/24, PC
(C12N15/09,
PC C12R1:07), (C12N1/21, C12R1:19), (C12N9/24, C12R1:19): CC
strandedness: Double;
CC topology: Linear;
FH key Location/Qualifiers
FH source 1. 1365
FT /organism='Bacillus sp.'
FT /strain='V230'
FT mat_peptide 1. 1365
FT /product='beta-fructofuranosidase'.
FEATURES
source
location/Qualifiers
BASE COUNT 459 a 266 c 283 g 357 t
ORIGIN
Query Match 100.0%; Score 1365; DB 6; Length 1365;
Best local Similarity 100.0%; Pred. No. 1.7e-305;
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGACACAGGGGAGCTACAGAGACATATGTTTCCCATATTCACGCGCTGACATG 60
DB 1 ATGACACAGGGGAGCTACAGAGACATATGTTTCCCATATTCACGCGCTGACATG 60
OY 61 CTAATAATTCAGAGACACAAACAGTCTCAATTAAGTCCCAATTCATGATGATCA 120
DB 61 CTAATAATTCAGAGACACAAACAGTCTCAATTAAGTCCCAATTCATGATGATCA 120
OY 121 GCAATCAAAAACATTTGATTTGGCAAAAAGGATGATTAAGTCAGGCACTTAATGATTTA 180
DB 121 GCAATCAAAAACATTTGATTTGGCAAAAAGGATGATTAAGTCAGGCACTTAATGATTTA 180
OY 181 GATGTATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 181 GATGTATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
OY 241 TATCATCATGCTCCGCTTTAGCAGGTGACCCCAAAAACAGATGATGATGATGATGAT 300
DB 241 TATCATCATGCTCCGCTTTAGCAGGTGACCCCAAAAACAGATGATGATGATGATGAT 300
OY 301 TTATCTATCAAAAAGTGGTATACATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 TTATCTATCAAAAAGTGGTATACATGATGATGATGATGATGATGATGATGATGATGAT 360
OY 361 TTTGAAGATATGATTAATTTGTTCCTAATGATCCGTATCTTAATATCAACACAGAG 420
DB 361 TTTGAAGATATGATTAATTTGTTCCTAATGATCCGTATCTTAATATCAACACAGAG 420
OY 421 TGTGAGGTCTGCTACTTAAACCAAGATGCGGAGCTGCTTTCATGATCAGATTAC 480

```





QY 481 TCAGGTAATCCTGAGATGTGGAGCCGGTCTGGTAACCAATCATTTCAACTGCTCAA 540  
Db TCAGGTAATCCTGAGATGTGGAGCCGGTCTGGTAACCAATCATTTCAACTGCTCAA 540  
QY 937 TCAGGTAATCCTGAGATGTGGAGCCGGTCTGGTAACCAATCATTTCAACTGCTCAA 540  
Db 541 GTAACCTATCCAGCCGGATGACACTTAAAGTCGATGGATCTGATATATA 600  
997 GTAACCTATCCAGCCGGATGACACTTAAAGTCGATGGATCTGATATATA 1056  
QY 601 TCTGCTTTGATGGCGGAGCGGTACAGTTATCAAAATATTACCAATTTATGATATA 660  
Db 1057 TCTGCTTTGATGGCGGAGCGGTACAGTTATCAAAATATTACCAATTTATGATATA 1116  
QY 661 GCGAAGTGATTTGAGGATTAACCATCTTTAAGAGACCTCACTATGTGAAGATAG 720  
Db 1117 GCGAAGTGATTTGAGGATTAACCATCTTTAAGAGACCTCACTATGTGAAGATAG 1176  
QY 721 GCGCATTAATATCTTGTCTTTGAGCGAATCTGGAACAACAGATGTTATCAAGCGAT 780  
Db 1177 GCGCATTAATATCTTGTCTTTGAGCGAATCTGGAACAACAGATGTTATCAAGCGAT 1236  
QY 781 CAGTCTTCATTAATTAAGTCTTACTATGCGGAGTACGTCCTTCTCCAGATGATAA 840  
Db 1237 CAGTCTTCATTAATTAAGTCTTACTATGCGGAGTACGTCCTTCTCCAGATGATAA 1296  
QY 841 AATAAAGCTGCTCAAGTCTTAATAAACAATTTGCTTTAGCGAATGCTGCATTAGC 900  
Db 1297 AATAAAGCTGCTCAAGTCTTAATAAACAATTTGCTTTAGCGAATGCTGCATTAGC 1356  
QY 901 ATTGTTGAATTTGGCGGATGATATATGATGAAATGTTATGAACCATTTAGTCATCA 960  
Db 1357 ATTGTTGAATTTGGCGGATGATATATGATGAAATGTTATGAACCATTTAGTCATCA 1416  
QY 961 AACACAGTAGAGATGAAGTGAAGCGCGCAATATATTAAATGAATTAATGAATGAT 1020  
Db 1417 AACACAGTAGAGATGAAGTGAAGCGCGCAATATATTAAATGAATTAATGAATGAT 1476  
QY 1021 CTATTCACGAGTTCAGAGATTCCAAAATGACAGTGTGAATTAACGACAAGATGTT 1080  
Db 1477 CTATTCACGAGTTCAGAGATTCCAAAATGACAGTGTGAATTAACGACAAGATGTT 1536  
QY 1081 TATATGCTAAGGCGCGGAGGAGCTCTTAATATGCGCCACACACACCGCATTAATGA 1140  
Db 1537 TATATGCTAAGGCGCGGAGGAGCTCTTAATATGCGCCACACACACCGCATTAATGA 1596  
QY 1141 GGACTTTATTTGAACATGAATCTTGACCTGCTGATCTCACACACTTACTCTCATTTG 1200  
Db 1597 GGACTTTATTTGAACATGAATCTTGACCTGCTGATCTCACACACTTACTCTCATTTG 1656  
QY 1201 GGTATCCGCAACCTGAAGGTAATATGTGTAATCACAAGTTATATGACGAATAGAGC 1260  
Db 1657 GGTATCCGCAACCTGAAGGTAATATGTGTAATCACAAGTTATATGACGAATAGAGC 1716  
QY 1261 TTTATCCAGAAACATCTCTACCTGGGGAACAAGCTTTGGGTTAATTAAGGGTCT 1320  
Db 1717 TTTATCCAGAAACATCTCTCTACCTGGGGAACAAGCTTTGGGTTAATTAAGGGTCT 1776  
QY 1321 GACACATCTGGAGAGAAATAGTCCGAGACAAGACATTTCCCA 1365  
Db 1777 GACACATCTGGAGAGAAATAGTCCGAGACAAGACATTTCCCA 1821

RESULT 5  
AR078182 2408 bp DNA linear PAT 31-AUG-2000  
LOCUS AR078182 Sequence 5 from patent US 5962297.  
ACCESSION AR078182  
VERSION AR078182.1 GI:10004928  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2408)  
AUTHORS Tsusaki,K., Kubota,M. and Chaen,H.

TITLE Polypeptides having .beta.-fructofuranosidase activity  
JOURNAL Patent: US 5962297-A 5 05-Oct-1999;  
FEATURES Location/Qualifiers  
source 1. 2408  
BASE COUNT 777 a 492 c 496 g 643 t  
ORIGIN

Query Match 100.0% Score 1365; DB 6; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 1.6e-305;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACACCGGGGATCTACAGAGACATATGTTTGGCCATATATACAGCGCTGATG 60  
Db 457 ATGAACACCGGGGATCTACAGAGACATATGTTTGGCCATATATACAGCGCTGATG 516  
QY 61 CTAAATTTCCAGACACAACAAGTCTCAATTTAAAGTGCCTCAATTCATATGATCA 120  
Db 517 CTAAATTTCCAGACACAACAAGTCTCAATTTAAAGTGCCTCAATTCATATGATCA 576  
QY 121 GCAATCAAAACATGATTCGGCAAAAGGATATGATGATGATGATGATGATGATGAT 180  
Db 577 GCAATCAAAACATGATTCGGCAAAAGGATATGATGATGATGATGATGATGATGATG 636  
QY 181 GATGTATGGGATGAGCTGGCCACTGCAAAACGCTGATGATGATGATGATGATGATG 240  
Db 637 GATGTATGGGATGAGCTGGCCACTGCAAAACGCTGATGATGATGATGATGATGATG 696  
QY 241 TATCATGCTGCTCCGCTTTTACAGGTGACCCAAAACAGTATGATGATGATGATG 300  
Db 697 TATCATGCTGCTCCGCTTTTACAGGTGACCCAAAACAGTATGATGATGATGATGATG 756  
QY 301 TTATCTCTCAAAAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATG 360  
Db 757 TTATCTCTCAAAAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATG 816  
QY 361 TTTGAAGATATGATGAATTTGTTCCAAATGATCCGATCTTAATATATCAAAACAG 420  
Db 817 TTTGAAGATATGATGAATTTGTTCCAAATGATCCGATCTTAATATATCAAAACAG 876  
QY 421 TGGTACAGTCTGCTACTTTTAACCAAGATGCGCAAGTCCGTTTATTCATACATTA 480  
Db 877 TGGTACAGTCTGCTACTTTTAACCAAGATGCGCAAGTCCGTTTATTCATACATTA 936  
QY 481 TCAGGTAATCCTGAGATGTGGAGCCGGTCTGGTAACCAATCATTTCAACTGCTCAA 540  
Db 937 TCAGGTAATCCTGAGATGTGGAGCCGGTCTGGTAACCAATCATTTCAACTGCTCAA 996  
QY 541 GTAACCTATCCAGCCGGATGACACTTAAAGTCGATGGATCTGATATATA 600  
Db 997 GTAACCTATCCAGCCGGATGACACTTAAAGTCGATGGATCTGATATATA 1056  
QY 601 TCTGCTTTGATGGCGGAGCGGTACAGTTATCAAAATATTACCAATTTATGATATA 660  
Db 1057 TCTGCTTTGATGGCGGAGCGGTACAGTTATCAAAATATTACCAATTTATGATATA 1116  
QY 661 GCGAAGTGATTTGAGGATTAACCATCTTTAAGAGACCTCACTATGTGAAGATAG 720  
Db 1117 GCGAAGTGATTTGAGGATTAACCATCTTTAAGAGACCTCACTATGTGAAGATAG 1176  
QY 721 GCGCATTAATATCTTGTCTTTGAGCGAATCTGGAACAACAGATGTTATCAAGCGAT 780  
Db 1177 GCGCATTAATATCTTGTCTTTGAGCGAATCTGGAACAACAGATGTTATCAAGCGAT 1236  
QY 781 CAGTCTTCATTAATTAAGTCTTACTATGCGGAGTACGTCCTTCTCCAGATGATAA 840  
Db 1237 CAGTCTTCATTAATTAAGTCTTACTATGCGGAGTACGTCCTTCTCCAGATGATAA 1296  
QY 841 AATAAAGCTGCTCAAGTCTTAATAAACAATTTGCTTTAGCGAATGCTGCATTAGC 900  
Db 1297 AATAAAGCTGCTCAAGTCTTAATAAACAATTTGCTTTAGCGAATGCTGCATTAGC 1356  
QY 901 ATTGTTGAATTTGGCGGATGATATATGATGAAATGTTATGAACCATTTAGTCATCA 960

Db	1357	ATTGTTGATTTGGCCGATGACTATACAGTGAAAAAGTGTATGAACCATTAAGTGCATCA	1416
QY	961	AACACAGTAGCAGATGAAGTCGAACGGCCCAATATATTTAAATGAATAAATAAAGGTAT	1020
Db	1417	AACACAGTAGCAGATGAAGTCGAACGGCCCAATATATTTAAATGAATAAATAAAGGTAT	1476
QY	1021	CTATTACAGGATTTAAAGAGATCCAAAATGACAGAGTGTGATGATTAACGACAAAGATGT	1080
Db	1477	CTATTACAGGATTTAAAGAGATCCAAAATGACAGAGTGTGATGATTAACGACAAAGATGT	1536
QY	1081	TATATGCTAGAGGGCCGGAGGCGACTCCCTTAAATGAGCCACACACCCGATTAATGAAGCT	1140
Db	1537	TATATGCTAGAGGGCCGGAGGCGACTCCCTTAAATGAGCCACACACCCGATTAATGAAGCT	1596
QY	1141	GGACTTGATTTGAACATGAAATCTTGACCCCTGCTATCTCACACACACTTACTCTCATTCG	1200
Db	1597	GGACTTGATTTGAACATGAAATCTTGACCCCTGCTATCTCACACACACTTACTCTCATTCG	1656
QY	1201	GGTATCCCGCACCCCTGAAGGTAAATATATGTGTACTCACAAAGTTATATGACAAATAGAAGC	1260
Db	1657	GGTATCCCGCACCCCTGAAGGTAAATATATGTGTACTCACAAAGTTATATGACAAATAGAAGC	1716
QY	1261	TTTCTATTCGACAAATCTCAGCTCTCACCTCGGGACAAAGCTTGGGGTTAATATTAAGGCTCT	1320
Db	1717	TTTCTATTCGACAAATCTCAGCTCTCACCTCGGGACAAAGCTTGGGGTTAATATTAAGGCTCT	1776
QY	1321	GACACATCTGGAGGAGAAATAGTTGTCGGACAAAGCAATTTCCCA	1365
Db	1777	GACACATCTGGAGGAGAAATAGTTGTCGGACAAAGCAATTTCCCA	1821

[illegible]

D	697	TATACATCGCTCCGCTTTAGCAGGTGACCCAAAAACAGTATGATCTCCACTTCA	756
O	301	TTATTCTATCAAAAAGTCGGTATACATTCGATTGACAGCTGGAAAAATGCTGAAAGTA	360
D	757	TTATTCTATCAAAAAGTCGGTATACATTCGATTGACAGCTGGAAAAATGCTGAAAGTA	816
O	361	TTTGAAGTATGATAAATTGTGTCACAAATGATCCGATCTTAAATATCAAAACAGGAG	420
D	817	TTTGAAGTATGATAAATTGTGTCACAAATGATCCGATCTTAAATATCAAAACAGGAG	876
O	421	TGTCAGGTTCTGCTACTTTAAACCAAGATGGCCAGTCGGTTATTCTATACAGATTAC	480
D	877	TGTCAGGTTCTGCTACTTTAAACCAAGATGGCCAGTCGGTTATTCTATACAGATTAC	936
O	481	TCAGTAATCCTGAAAGTGGTGGAAACCGGTGCTGTAAACCAATCTTCAACTGCTCA	540
D	937	TCAGTAATCCTGAAAGTGGTGGAAACCGGTGCTGTAAACCAATCTTCAACTGCTCA	996
O	541	GTAACCTTATCCACGCCGATGCACTGACATTAAGTCATGAGATATCATATAA	600
D	997	GTAACCTTATCCACGCCGATGCACTGACATTAAGTCATGAGATATCATATAA	1056
O	601	TCTGCTTTTGATGGCGGAGACGGTACGTTTATCAAAATTTACAGCAATTTATCGATGA	660
D	1057	TCTGCTTTTGATGGCGGAGACGGTACGTTTATCAAAATTTACAGCAATTTATCGATGA	1116
O	661	GGCAAGTGAATTTAGGTGATTAACATACCTTTTAAAGACACCTCAGTATGTTGAAGATAG	720
D	1117	GGCAAGTGAATTTAGGTGATTAACATACCTTTTAAAGACACCTCAGTATGTTGAAGATAG	1176
O	721	GGCCATAAATATCTTGTCTTTGAAGCGCAATACAGCAACAGATGGTTATCAAGCGAT	780
D	1177	GGCCATAAATATCTTGTCTTTGAAGCGCAATACAGCAACAGATGGTTATCAAGCGAT	1236
O	781	CAGCTTTTCATATATAAGCTTACTATGGGGGAAAGTAGCTCTTCCGAGATGAAAA	840
D	1237	CAGCTTTTCATATATAAGCTTACTATGGGGGAAAGTAGCTCTTCTTCCGAGATGAAAA	1296
O	841	AATAAAGTCTTCAAAAGTCTTAAAAAACAATTTGCTTCTTTAGCGAATGGTGCAATTAAGC	900
D	1297	AATAAAGTCTTCAAAAGTCTTAAAAAACAATTTGCTTCTTTAGCGAATGGTGCAATTAAGC	1356
O	901	ATTGTGTAATTTGGCCGATGACTATACAGTGTAAAAAGTATTGAAACCATTAAGTCGATCA	960
D	1357	ATTGTGTAATTTGGCCGATGACTATACAGTGTAAAAAGTATTGAAACCATTAAGTCGATCA	1416
O	961	AACACAGTAGCAGATGAAGTCGAACGGCCCAATATATTAAAAATGAATATAATGCTAT	1020
D	1417	AACACAGTAGCAGATGAAGTCGAACGGCCCAATATATTAAAAATGAATATAATGCTAT	1476
O	1021	CTATTTCACGATTCAGAGATCCAAAAATGACAGATGATGAATTAACGACAAAGATGTT	1080
D	1477	CTATTTCACGATTCAGAGATCCAAAAATGACAGATGATGAATTAACGACAAAGATGTT	1536
O	1081	TATATGCTAGGAGCCCGAGGAGACTCTTAAATGCGCCACACAAACCGATTAATGAACCT	1140
D	1537	TATATGCTAGGAGCCCGAGGAGACTCTTAAATGCGCCACACAAACCGATTAATGAACCT	1596
O	1141	GGACTTTGATTTGAACATGAATCTTTGACCCCTGCTATCTCACACACACTTACTCTCATTTG	1200
D	1597	GGACTTTGATTTGAACATGAATCTTTGACCCCTGCTATCTCACACACACTTACTCTCATTTG	1656
O	1201	GGTATCCCGACACCTCGAAGGTAAATATGTGTACTCTCAAGTTATATAGACAAATAGAGC	1260
D	1657	GGTATCCCGACACCTCGAAGGTAAATATGTGTACTCTCAAGTTATATAGACAAATAGAGC	1716
O	1261	TTTCTATCCAGAACATCACTCTCACTCTCGGGACACAGTTGGGTTTAAATTAAGGCTCT	1320
D	1717	TTTCTATCCAGAACATCACTCTCACTCTCGGGACACAGTTGGGTTTAAATTAAGGCTCT	1776
O	1321	GACACATCTGGAGAGAAATAGTTCCGACCAAGACAAATTTCCCA	1365

Db 1777 GACACATCTGAGAGAGAAATAGTTCCGGACAAGACAATTCCCA 1821

RESULT	7
LOCUS	E15383
DEFINITION	2408 bp DNA linear PAT 28-JUL-1999
ACCESSION	E15383
VERSION	E15383.1 GI:5710066
KEYWORDS	JP 199806586-A/2.
SOURCE	Bacillus sp.

Query Match	100.0%	Score 1365;	DB 6;	Length 2408;
Best Local Similarity	100.0%;	Pred. No. 1,66-305;		
Matches 1365;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	ATGAAACAGCGGGGACTACAGGAAGACTATGCTTTTGCCCATATTATACAGCGCTGCATCG	60
Db	457	ATGAACAGCGGGGACTACAGGAAGACTATGCTTTTGCCCATATTATACAGCGCTGCATCG	516
QY	61	CTAAAAATTCAGAGACAACAAAACAGTCCTCAATTAAAGCGCCCAATTCATAGCATCA	120
Db	517	CTAAAAATTCAGAGACAACAAAACAGTCCTCAATTAAAGCGCCCAATTCATAGCATCA	576
QY	121	GCAATCAAAAAACATTGATTGCGCAAAAGGGATGATTAAGTCAGGCAACTTAATAGATTTA	180
Db	577	GCAATCAAAAAACATTGATTGCGCAAAAGGGATGATTAAGTCAGGCAACTTAATAGATTTA	636
QY	181	GATGTATGGGTAGCTGGCCACTGCCAAAAGCGCTATGCTACGGCGCAAAATTATCATGGA	240
Db	637	GATGTATGGGTAGCTGGCCACTGCCAAAAGCGCTATGCTACGGCGCAAAATTATCATGGA	696
QY	241	TATACATCGCTCCGCTTTAGCAGGTGACCCAAAAAACACTGATGATACCTCCACTTAT	300
Db	697	TATACATCGCTCCGCTTTAGCAGGTGACCCAAAAAACACTGATGATACCTCCACTTAT	756
QY	301	TTATTTCATCAAAAAGTCGATGATACATCGATTGACAGCTGGAAAAAATGCTGGAAGAGTA	360
Db	757	TTATTTCATCAAAAAGTCGATGATACATCGATTGACAGCTGGAAAAAATGCTGGAAGAGTA	816
QY	361	TTTGAAGATATGATTAATAATTTGTTCCAAATGATCCGTAATCTTAAATATCAACAACAGGAG	420

Db	817	TTTGAGAGTATGAGTAAATTTGTTCCAAATGATCCGTAATCTTAAATATATCAACACAGAG	876
QY	421	TGTCAGGTTCTGCTACTTTAAACCAAGATGGCAAGTCGTTTATCTATACAGATTAC	480
Db	877	TGTCAGGTTCTGCTACTTTAAACCAAGATGGCAAGTCGTTTATCTATACAGATTAC	936
QY	481	TCAGTAATCCTGSAAGATGNTGSAACCGGTCCTGTATACCAAAATCATTTCAACGCTCAA	540
Db	937	TCAGTAATCCTGSAAGATGNTGSAACCGGTCCTGTATACCAAAATCATTTCAACGCTCAA	996
QY	541	GTAACCTATCCACGCGGATGACAGCTTCACGTTAAAGTCGATGAGATTCGATCATPAA	600
Db	997	GTAACCTATCCACGCGGATGACAGCTTCACGTTAAAGTCGATGAGATTCGATCATPAA	1056
QY	601	TCTGCTTTGATGGCGAGACGGTACGTTATCAAAATATTCAGCAATTTATCGANGAA	660
Db	1057	TCTGCTTTGATGGCGAGACGGTACGTTATCAAAATATTCAGCAATTTATCGANGAA	1116
QY	661	GGCAAGTGGATTTACAGTGATACCATCTTAAAGACCCCTGATAGTGGAGATPAG	720
Db	1117	GGCAAGTGGATTTACAGTGATACCATCTTAAAGACCCCTGATAGTGGAGATPAG	1176
QY	721	GGCCATAAATATCTTGTCTTTGAACGGAATCTGGACACAGATGTTATCAAGCGAT	780
Db	1177	GGCCATAAATATCTTGTCTTTGAACGGAATCTGGACACAGATGTTATCAAGCGAT	1236
QY	781	CAGTCCTTCATATATAAGCTTACTATGCGGGAAGTGAGCTCTTCCACAAATGAANA	840
Db	1237	CAGTCCTTCATATATAAGCTTACTATGCGGGAAGTGAGCTCTTCCACAAATGAANA	1296
QY	841	AATAAAGCTCTTCAAAAGTCTTAAAAAACAATTTGCTTTTACGGGAATGTGCAATTAGC	900
Db	1297	AATAAAGCTCTTCAAAAGTCTTAAAAAACAATTTGCTTTTACGGGAATGTGCAATTAGC	1356
QY	901	ATTGTGTGAATTTGGCCGATGACTATACAGTGAAGATGTTATGAACCATTTAGTCGATCA	960
Db	1357	ATTGTGTGAATTTGGCCGATGACTATACAGTGAAGATGTTATGAACCATTTAGTCGATCA	1416
QY	961	AACACAGTAGAGATGAAGTGAAGCGGCCCATATATTTAAATGAATAATTAATAGTAT	1020
Db	1417	AACACAGTAGAGATGAAGTGAAGCGGCCCATATATTTAAATGAATAATTAATAGTAT	1476
QY	1021	CTATTACAGATTCAGAGATCCAAAATGACGAGTGAATTAACGACAAAGATGT	1080
Db	1477	CTATTACAGATTCAGAGATCCAAAATGACGAGTGAATTAACGACAAAGATGT	1536
QY	1081	TATATGCTAGAGGCCCGGAGCGGACTCCCTTAATGGCCCAACACCCGATTAATGAAC	1140
Db	1537	TATATGCTAGAGGCCCGGAGCGGACTCCCTTAATGGCCCAACACCCGATTAATGAAC	1596
QY	1141	GGACTTGTATGAACATATATCTTGACCCCTGCTGATCTCACACACACTTACTCTCATTC	1200
Db	1597	GGACTTGTATGAACATATATCTTGACCCCTGCTGATCTCACACACACTTACTCTCATTC	1656
QY	1201	GCTATCCCGCACCCCTGAAGGTATTAATGTGTACTACAAATTTATATGACAATAGAGGC	1260
Db	1657	GCTATCCCGCACCCCTGAAGGTATTAATGTGTACTACAAATTTATATGACAATAGAGGC	1716
QY	1261	TTTCAATCCAGAAATCATCTTCACCTCGGGGACAAGCTTGGGTTAAATATTAAGGGCT	1320
Db	1717	TTTCAATCCAGAAATCATCTTCACCTCGGGGACAAGCTTGGGTTAAATATTAAGGGCT	1776
QY	1321	GACACATCTGGAGAGGAATAATAGTTCCGGACAGAGCAATTTCCA	1365
Db	1777	GACACATCTGGAGAGGAATAATAGTTCCGGACAGAGCAATTTCCA	1821

RESULT 8	
AE007686	
LOCUS	11318 bp
DEFINITION	DNA linear BCT 27-JUL-2001
	Clostridium acetobutylicum ATCC824 section 174 of 356 of the complete genome.

ACCESSION AE007686 AE001437  
 VERSION AE007686.1 GI:15024739  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Clostridium acetobutylicum.  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium  
 1 (bases 1 to 11318)  
 Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,O.,  
 Gibson,R., Lee,H.M., Dubois,J., Olu,D., Hitti,J., Wolf,Y.I.,  
 Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,  
 Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.  
 Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum  
 J. Bacteriol. 183 (16), 4823-4838 (2001)  
 21359325  
 11466286  
 2 (bases 1 to 11318)  
 Childress,D., Zeng,O. and Smith,D.R.  
 Direct Submission  
 Submitted (24-JUN-2001) GTC Sequencing Center Production,  
 Flushing and Bioinformatics teams, Genome Therapeutics Corp., 100  
 Beaver Street, Waltham, MA 02453-8443, USA  
 Location/Qualifiers  
 1. 11318  
 /organism="Clostridium acetobutylicum"  
 /strain="ATCC 824"  
 /db\_xref="ATCC:824"  
 /db\_xref="taxon:1488"  
 89. 790  
 /gene="CAC1770"  
 89. 790  
 /gene="CAC1770"  
 /codon\_start=1  
 /transl\_table=11  
 /product="RNA polymerase sigma factor, SigK-like"  
 /protein\_id="AAK79735.1"  
 /db\_xref="GI:15024740"  
 /translation="MLENRSLEDRLQKAKESKELNRLIOEKPEIASLQKTKGYL  
 OYGDDELSTGIMAFEEAVESYDRNGKELSPARKVITLRAIDYRKQKRNVIYTLA  
 GNVSDNDIDIMPIENFKAMEQYKNSKINERIRLELEPKRLREMGIEFVULVASPK  
 OEKLKTKYKIANVIVHMODILNKLIRKRLPIKEITEKNITINRKLKERGIYITALV  
 LAETGDRLEIREYLE"  
 801. 1544  
 /gene="CAC1771"  
 801. 1544  
 /gene="CAC1771"  
 /codon\_start=1  
 /transl\_table=11  
 /product="uncharacterized protein, ykri B subtilis  
 homolog"  
 /protein\_id="AAK79736.1"  
 /db\_xref="GI:15024741"  
 /translation="MKGIVLELKYKLIILTEKGFDEIDKLNTRAKIGDINIKSKK  
 ISTIKALREFLAATALLIFPILNSLIYGMFTREYVVDINPNNNKVAWEIHYNF  
 GNIRKLAANKGVSIIVKMRNLKFKATNVINNEIRTAQDKRIDSQKQNTYITTA  
 SSSSIDESIDSLERIKENKINARPMIVGNDIDYKSKQYGIPIDKRILLNKVI  
 KNPSTKPYDIDNKKSIDELINIANQENY"  
 1682. 3151  
 /gene="CAC1772"  
 /note="sacB"  
 1682. 3151  
 /gene="CAC1772"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Levanucrase"  
 /protein\_id="AAK79737.1"  
 /db\_xref="GI:15024742"  
 /translation="MNKLKIVKCLIGSMISGIIITOOTFASTNDMNKENTYGVSHIT  
 RYNSKIPMEQNDLKPRYPOBNASTIKINIASAKYDKNGNLDIDVDSMPLQWGDGT  
 VVANHGYHIFALAGDKPRNODDTIYMPYQKIGENSIDSKNACKVFRDSKIYANDP  
 YLKTQEMSSATLISDGOVRLFTDFSGVAKGQGDASQVITTTQVNLSDPSNT  
 INIDVSDHKSVPFGNGTIVYNOQFIDEGKSSGDNHTLRDPHYVEDNGRKYLVEE

AMGTNDGYGDTSLNKAFYGRSOSFEKTEKDOLLIDTNKKHDAISLANGALGIIEN  
 NDYTLKKEMKPLIASNTVTEIERANFKMGKMYLFTDSDGSKMTNGISSKQIYAL  
 GFSSNSL/GPKYKPLNKGVLNLDLPDLPFTYSHFAVPQTNKNKVVITSYITNRGA  
 YSDHSSFPAPSPFLINIKGTSTVSNISLQGGQLTIDNY"  
 3249. 4793  
 /gene="CAC1773"  
 /note="sacC"  
 3249. 4793  
 /gene="CAC1773"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Levanase"  
 /protein\_id="AAK79738.1"  
 /db\_xref="GI:15024743"  
 /translation="MKRRKLKILFIIEFYIILISLYIONNSTIKSTTKKONY  
 REYHPTVPNKKNDPDRPVPPEGEYHYIYKYNADPNNGNTEMRQATSDIYWRDE  
 GCIPKPTTNGDITWSSGFVYDAONTGFEKGALVALVITPSPASMDQEOFLYSTR  
 GRTKPYSNQIIPNCTKDFRDKIIMDERKNKMWVLAEGTIGFYESTNLKNQW  
 TGDFETNNIGIVECPDIEMQSDNGNKKWILGTSANCKVSGEPYVAVWNGDGGKE  
 IADISTPKMLDYGEDWYAAVTESENSKNLKKRYALAMNNMDYANNPTIQRNPG  
 MDSIVRETTLSKQKDNNTYSLSKRIKIENTTISIDQKQISYGLKHLKVGKAYOL  
 DTDINWDARKWGIRLNESLDRKHHDVGFTEBKYSYVRATYGNPDGSKKYYESA  
 PFDINNKVHLIRFVDRVSEVFIIDGRITYSNEVFPREDKGITLFSIRKAVFKNI  
 TIKHID"  
 4860. 6146  
 /gene="CAC1774"  
 /note="sacB2"  
 4860. 6146  
 /gene="CAC1774"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Levanucrase"  
 /protein\_id="AAK79739.1"  
 /db\_xref="GI:15024744"  
 /translation="MKTRTKYKMISSMLIALITTEPLLIRHNTGYTSISROAON  
 FKCTRENTAPINPNEFKLITAPNLWMDTPVLRKQDGLAVANGIKVIFALPASNVGM  
 NKRRDVAGISYFCSTGDEBNVYKGLAVNEBALGSRQWASALIDENGWQOEFYATG  
 RKGEAVTFEQRLLTKTFSLINVDKGVHITCSKHQVILPEPDGYVYOTMOAKPITV  
 SERPVEFEDPKTKDYLIFEGNKGKRIEKKKPNIDGKLEPKHNIAPRGVENGVAV  
 GIVAAOKMDLRFRLPLLELAVGVNOOLERPOIWKVKNKYILFTISHKFTYAGGLNG  
 VDGIHGCGNSLRBNRYPLNGVAVITNPTRDPQYTSWLVSGHDVLSFTNEHPNG  
 QLRYGTFAPLQIISLKGYSKTIIGLGEGVYTPAH"  
 6560. 8305  
 /gene="CAC1775"  
 6560. 8305  
 /gene="CAC1775"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Predicted membrane protein"  
 /protein\_id="AAK79740.1"  
 /db\_xref="GI:15024745"  
 /translation="MKIKGIIAISMSLITGSISSVYKASSNNKRLONTYKEROKIL  
 LNKDSNKSIVTNDNNEKLSLDDLNKKNKVIPEFNKTIILNKKIKLPIITSGDYSKTK  
 GNSVASAGDSNNAAPAKLDPIYILNBSLKNGBITNTNTOJAWISDPTDADGVLSTA  
 VGGPSPAYIILGLDNDNGPVYQFNPAGDFDLYVOTPSKGAKNIVGSIHVIPEDY  
 OAPGTFQSTSTYVNAVAVDESKMDTAICLTGTSQDMNLKISDQSGVAVQVNVY  
 DKWVYINKPSNAGCINNTITTPYVSSNSYSLRMGNKNDVEAMLSGIONTI  
 ILDKTYENNKQFOETPTPKNOEMEPETADGSVITLLNNYPELRFQVPTDELOYL  
 FTDNPNNASTHRSQYCGLNKTKAEKRLNTAGODYLVLYISPLIOTDPIILNDEIH  
 LTVGDPAVLIYDSTTAYADNYITGPTTYSSIANINVTNVPTAAVKDIYASLNDVAR  
 VSDENYKVMAPGESFWRDSGOQCINIGIDVNVYENGNNVQLMGTWRFVQAGVDPIS  
 IIPGLHFDYVEAGD"  
 8693. 9346  
 /gene="CAC1776"  
 8693. 9346  
 /gene="CAC1776"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HAD superfamily hydrolase"  
 /protein\_id="AAK79741.1"  
 /db\_xref="GI:15024746"  
 /translation="MDRIITFEDCYGLIDTRKPVFDAISNIAKNNILNSKDAVINIYVY  
 EDRLMVGETIYIPDKIKQVLEYCDMELVTNVFSSYSDEIISVHKRLKPKREVESLE

KHKYELALMSVHDIIDYNDALGNFNNVLSEDIHAYPQLEKFEYVERLE  
LKORHCHIAQWMDIIPCSKIGMRKIWRKNGKNKCHEPEVEVHTDEVINLI\*  
/gene=.10521  
/gene="CAC1777"  
9598..10521  
/gene="CAC1777"  
/codon\_start=1  
/transl\_table=1  
/product="Nudix (Mult-like) hydrolase"  
/protein\_id="AAK79742.1"  
/db\_xref="GI:15024747"  
/translation="MDMNKKNRNGLTKEKFLDRVPDGERPSNTVMDLFTVDVY  
IEGDPNKAALKILIKRGDHPYMGCAVPGGFVNIINGLSACRYRELKEFNVNYF  
EOLTFPDVDRMREIVSYAMALADKISITPKAGDDADADAFYVKKFISKGF  
GIDKIDYNIILTDDEETKIGYLVTEKEKNGYVITIKVPSRLLEMSKEELADHIE  
EIDCALRIRKIKETYPPIASRLPKYFTLIRPAKVFAIILNLEKPLIRANRRRIKKM  
VVEREKRTISGRATCTYTNEDMOTFLDE"  
10543..11193  
/gene="CAC1778"  
10543..11193  
/gene="CAC1778"  
/codon\_start=1  
/transl\_table=1  
/product="Amidase from nicotianamide family"  
/protein\_id="AAK79743.1"  
/db\_xref="GI:15024748"  
/translation="MGKFEKILNMSEIKRFPVNDLDDKDTAIVVDVNGFVHE  
GALSSPVEGIVDEIVRIKNEKTKKIFLDEHTNSTEKSYAKHLESLAEELI  
PELKNELDSNTYMIPIKNSYNGFHAFGRKMKLEPNESQIENYICCGEDICSNFA  
NTLKTYNOKMMDRIIIPSNVETFDGHDLMKILISLWEMOSNGIETVDRVL"  
BASE COUNT 4466 a 1373 c 1945 g 3534 t  
ORIGIN  
Query Match 55.4% Score 755.8; DB 1; Length 11318;  
Best Local Similarity 73.4%; Pred. No. 1.4e-164;  
Matches 980; Conservative 0; Mismatches 352; Indels 3; Gaps 1;  
QY 12 GGACTACAGGAGACATGATGTTTGGCCATATTACACGCGCTGACATGCTAAATAATCC 71  
DB 1777 GAATTACAAAGAAACATATGAGTTTGCATATTACCTGTTATATATGCTCAAAAATACC 1836  
QY 72 AGGACACAAAACGCTCTCAATTTAAAGTCCCTCAATTCATGATGATGCAATCAAAA 131  
DB 1837 AATGACACAAATATCTTAAATTTAAAGTCCCTCAATTTATCATCATCAATTTAAAAA 1896  
QY 132 CATGATTCGGCAAAAGGATGATGATGACGCACTTATGATTTAGATTTGATGGA 191  
DB 1897 TATAGCTTCAGCAAAAGGTTTATGATAGCAATGTTATTTGATTTGATGGA 1956  
QY 192 TAGCTGGCCACTGCAAAAACGCTGATGCTACGCAAAATTTATGATGATTCACATCGT 251  
DB 1957 TAGCTGGCCATTTGCAAAATGATGATGGAACAGTGGCTAACATGCTGATTCATATGCT 2016  
QY 252 CTCCGCTTTAGCAGTACCCAAAACAGATGATGATCTCCACTTCATTTATTCATCA 311  
DB 2017 TTTTGGCTTTGGCAGTATCTTAAATCAAGATGATCTCTATTTATGATTTATCA 2076  
QY 312 AAAAGTGGTATCATGATGATGACAGCTGAAAAATCTGGAAGATATTGGAAGATAT 371  
DB 2077 AAAAGTTGGAGAAAATCTCATTTGATTTGAGAGATGCAAGGTAAAGTTTAAAGATAG 2136  
QY 372 GGATTAATTTGTTCAATGATCGGTATCTTAAATATCAACAGAGAGTGTGAGTTC 431  
DB 2137 TGATTAATTTAGCTTCATGATCTTATCTTAAGATCAACACAGAGATGTGTGCTC 2196  
QY 432 TGTACTTTTAAACCAAGATGCGCAAGTCCGTTTATTCATACAGATTTACTAGTATCC 491  
DB 2197 TGCACCTTTTAAACATCTGATGCTCAAGTTCCTTTTATTCACAGATTTTTCGGAGTGC 2256  
QY 492 TGAAGATGTTGGAACCGGTGCTGGAACCAATCATTTCACTGCTGCAAGAACTTATC 551  
DB 2257 AAAAGATGTTGGAACAGATGCTAGTATCAAGTTATTAACAACACTGCTCAAGGAATTATC 2316

QY 552 CCAGCCGATCAGCTACACTTAAAGTGATGAGTATGATCATTAATCTGTTGA 611  
DB 2317 TCAACCAATTTCAATATTAATTAATGATAGTGAAGTGAATCATTAATCTGTTTGA 2376  
QY 612 TGGCGAGACGCTACAGTTTATCAAAATATTACAGATTTATTCATGAGGCAAGTGCAT 671  
DB 2377 CGGAGGAAATGGGCAATCTATCAGAACGTTCCAGCAATTTATGATGAAAGTGAATGAG 2436  
QY 672 TTCAGGCTATACCTACTTAAAGACCTCAGATCTGGAAGTGAAGGCTTGAATA 731  
DB 2437 TTCTGGCATATATCATCTACTTACAGATCTCATATTATGTAAGATTAATGGCCCAATA 2496  
QY 732 TCTTCTCTTTGAAGGAATACCTGGAACAACAGATGTTATCAAGCGATCAGTCTTCA 791  
DB 2497 TCTTGTTTTCGAGGTACACAGAGAACAGATGTTATGCAAGAGATFACCTTCTACT 2556  
QY 792 TAATTAACCTTACTATGCGGAGTGCAGCTCTTCTCCAGATGAAAAATAA---ACT 848  
DB 2557 CAATTAACCTTTTATGTTAGTAAGTCAAGTCTTCTTAAACAGAAAAAGATCAATTAAT 2616  
QY 849 GCTTCAAGTCCCTAATAAACAATTTGCTTTAGCAAGTGTGATTAAGCATTTGTA 908  
DB 2617 AATTGATCTATTAATAAAGCATGATGCTTTCTTTAGCTATGCGGCACTAGCATTAATGA 2676  
QY 909 ATTGCGCGATGATATACATGATGAAAGTGTATGAACCATTAATGCTGATCAACACAGT 968  
DB 2677 ATTAATATATGCTATATATTAATAAAGCATGATGCTTTCTTTAGCTATGCGGCACTAGCATTAATG 2736  
QY 969 AGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1028  
DB 2737 AACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2796  
QY 1029 GATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1088  
DB 2797 AGATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2856  
QY 1089 AGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1148  
DB 2857 AGGTTTGTGTTCAATGCTTGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2916  
QY 1149 ATTGACATGATCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1208  
DB 2917 ACTAATTTTAAACCTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2976  
QY 1209 GACCCGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1268  
DB 2977 TCAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3036  
QY 1269 AGAATCATCTCTCCTGCGGAGCAACCTTGGGTTAATTAAGGCTGTCACATC 1328  
DB 3037 AGATCATCATCTGATGTTTGCACCAAGTTTCTACTTAATTAAGGAGCAATGATTC 3096  
QY 1329 TGGAGGAGAAATAG 1343  
DB 3097 TGTATATCAAAATAG 3111  
RESULT 9  
AX434229  
LOCUS AX434229 1449 bp DNA linear PAT 28-JUN-2002  
DEFINITION Sequence 2644 from Patent WO0229113.  
ACCESSION AX434229  
VERSION AX434229.1 GI:21659037  
KEYWORDS  
SOURCE  
ORGANISM Bacillus licheniformis.  
Bacillus licheniformis  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
REFERENCE  
1 Berka, R. and Clausen, I. G.  
METHODS for monitoring multiple gene expression  
Patent: WO 0229113-A 2644 11-APR-2002;  
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)  
Location/Qualifiers



```

/transl_table=1
/protein_id="CA26512.1"
/db_xref="GI:40119"
/db_xref="SPTREMBL:O45663"
/translation="MCKGQSYFSGVTPYIF"
repeat_unit
337..370
446..455
/note="Imp. Inverted repeat A'"
RBS
464..1885
/note="pot. rRNA binding site (sacB)"
CDS
464..1885
/codon_start=1
/transl_table=1
/product="Levansucrase"
/protein_id="CA26513.1"
/db_xref="GI:732568"
/db_xref="SWISS-PROT:P05655"
/translation="MNIRKRRKQATVLTFTTALLAGTAQAFKETNPKYETGIS
HTRHMDLQIPEDQKNEKYDPEFDSSTIKNISAKGLDWDSPMLNADSTVANYH
YHIVFALAGDEPKNADPISLYMFKYQVETSIDSWKNAGRVKDSKPDANDSTLKDQ
OEWGSAFTSDGKIRLPTDFSGHYKQTLTAAVNVSAJSDSLNNGEDYKSIIF
DGDGKYONVOQFIDEENYSSGDNHLLRDPHYVEDGKHXYLFEANTGETEDYGEIS
LKNAYYKSTSPFROESOKLSDKRRPELALGACMTEINDYTLKMKRKLIS
NWTYDELRANVEKMGKWLFTDSRGSKMTIDGITSNDIYMLCYVNSLTPKPLN
KTLGLVLMDDPNDVPTFTYSHFAVPAQKGNVVTISYMTNKGFYADKQSTFAPSFLN
IKGRKTSVVRKDSIDEGQLTVNK"
slg_peptide
464..550
mat_peptide
551..1882
/product="mature Levansucrase (aa 1-444)"
misc_feature
1892..1928
/note="region of dyad symmetry"
repeat_unit
1892..1904
/note="inverted repeat B"
repeat_unit
1916..1928
/note="inverted repeat B'"
BASE COUNT
720 a 422 c 370 g 495 t
ORIGIN
Query Match 42.4% Score 578.6; DB 1; Length 2007;
Best Local Similarity 67.2%; Pred. No. 1.6e-123;
Matches 903; Conservative 0; Mismatches 389; Indels 51; Gaps 4;
QY 1 ATGACACGGGGGCTACAGAGACTGTTGGCCATATTACACGGCGTACATG 60
DB 557 ACGAACCAAAACATATAGGAAACATACGGCATTTCCCATATTACACGCATATGAT 616
QY 61 CTAAAAATTCAGGACAAACAGCTCAATTTAAAGCCCTCAATTCATAGCAATCA 120
DB 617 CTGCAAAATCCCTGACACGCAAAATAATGAAATATCAAGTCCGAAATTCGATTCG 676
QY 121 GCAATCAAAACATTTGATTCGGCAAAAGGTATGATAAGTCAGCAACTATATGATT 180
DB 677 ACAATTTAAATAATCTCTCTGCAAAAGG-----CCGTG 709
QY 181 GATATATGATAGCTGCGCACCTGCAAAACGCTGATGCTACTGCGCAATATATCANGA 240
DB 710 GACCTTTGGACACCTGCGCATTTACAAAACGCTGACGCACTGCGCAAACTATCAGCG 769
QY 241 TATACATCGTCGCTTGAAGGAGTGACCCAAACAAACATGATGATCTCCACTTCAT 300
DB 770 TACACATCGTCTTTGATTAAGCCGAGATCTTAAATGCGATGACACATCATTTAC 829
QY 301 TTATTCATCAAAAGTCGTATGATCATGATGACAGCTGGAATAATGTCGAAGAGTA 360
DB 830 ATGTTTATCAAAAGTCGGGGAACCTTCTATGACAGCTGGAATAACGCTGCGCGCTC 889
QY 361 TTGACATATGATTAATTTGTTCCAAATGATCGTATCTTAAATATCAACACAGAG 420
DB 890 TTTTAAACACGCGCAAAATGATGATGATCTTATCTTAAAGACCAACCAACAA 949
QY 421 TGTGAGGTTCTGCTACTTAAACCAAGATGCGCAAGTCCGTTATTTATATACGATTAC 480
DB 950 TGTGAGGTTCTGCGCAATTTACATCTGACGGAATAATCCGTTATTTCTACATGATTTC 1009

```

```

QY 481 TCAGGTATCCTGAAGATGTTGGAACCGGTCTGTGTAACCAATATTTCACTGCTCAA 540
DB 1010 TCCGGTAACATTTA-----CCGCAACAAACAGCTGACACGCAACAA 1051
QY 541 GTAACTTATCCACGCGGATGACACTTAAGTGCATGATGATGATGATGATGATGAT 600
DB 1052 GTTAACTGATCAGATCAGACAGCTCT---TTGAACATCAACGGGTGTGAGAGATTAT 1108
QY 601 TCTGCTTTGANTGGCGGAGGATGACATGATTAATCAAAATATTCAGCAATTTATGATGAA 660
DB 1109 TCAATCTTTGA---CGGTGACGGAATAAGATGATGATGATGATGATGATGATGATGAA 1165
QY 661 GCGAAGTGTATTCAGGTATACCATCTTAAAGAGACCCCTCATATGTTGAAGATTAAG 720
DB 1166 GCGAATCATACCTGACGGGACACATGATGATGATGATGATGATGATGATGATGATGAT 1225
QY 721 GCGCATTAATATCTTTGTTGAACCGCATCTGGAACACATGATGATGATGATGATGAT 780
DB 1226 GCGCACAATATCTTATGTTGAAGCAACACCTGGAACATGATGATGATGATGATGATGAT 1285
QY 781 CAGTCTTCATTAATTAAGCTTACTATGCGGAGTGACGTCCTTCTCCAGATGATAAA 840
DB 1286 GAATCTTTATTTAAACAAAGCATACTATGCAAAAGCACATCATCTTCTCGTCAAGAAAGT 1345
QY 841 AATAAATCTGTTCAAAAGTCTTAAACAAATTCCTTTTAAAGCAATGATGATGATGATGAT 900
DB 1346 CAATAAATCTTCTGCAAAAGGATTAATAAAGCAACGCTGATGATGATGATGATGATGATGAT 1405
QY 901 ATTGTTGATTTGGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 1406 ATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1465
QY 961 AACACAGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 1466 AACACAGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1525
QY 1021 CTATTCACGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGAT 1080
DB 1526 CTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1585
QY 1081 TATATGCTAGGCGCCGAGGAGCTCTTAAATGCCCCACACAAACCGATTAATGAACAT 1140
DB 1586 TACATGCTAGGCTTATGATTTTAAATCTTAACTGCCCCATACAAACCGCTGATCAACAACT 1645
QY 1141 GGACTTGTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1646 GGCTTGTGTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1705
QY 1201 GGTATCCCGCACCTGAGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1706 GCTGATCTCAACGGAAGGAAGCAATCTGATGATGATGATGATGATGATGATGATGATGAT 1765
QY 1261 TTCTATCCAGAACATCTCTACCTGCGGGAACAGCTTGGGGTTAATATTAAAGGCTCT 1320
DB 1766 TTCTATCCAGAACATCTCTACCTGCGGGAACAGCTTGGGGTTAATATTAAAGGCTCT 1825
QY 1321 GACACATCTGAGAGGAGGAATAAG 1343
DB 1826 AAAACATCTGTGTTCAAAAGACAG 1848

```

RESULT 11  
 AF047518/c  
 LOCUS AF047518 5831 bp DNA circular SYN 22-JUL-1998  
 DEFINITION Cloning vector pEX18cm, complete sequence.  
 ACCESSION AF047518  
 VERSION AF047518.1 GI:2961140  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Cloning vector pEX18cm.  
 Cloning vector pEX18cm  
 artificial sequences; vectors.  
 REFERENCE  
 1 (bases 1 to 5831)  
 Hoang, T.T., Karkhoff-Schweitzer, R.R., Kutchma, A.J. and

TITLE	Switzerland, H. P.
JOURNAL	A broad-host-range FLP-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked <i>Pseudomonas aeruginosa</i> mutants
MEDLINE	Gene 212 (1), 77-86 (1998)
PUBMED	98326312
REFERENCE	9661666
AUTHORS	2 (bases 1 to 5831) Hoang, T. T., Karkhoff-Schweitzer, R. R., Kutchma, A. J. and Schweizer, H. P.
TITLE	Direct Submission
JOURNAL	Submitted (10-FEB-1998) Microbiology, Colorado State University, Fort Collins, CO 80523, USA
FEATURES	Location/Qualifiers
SOURCE	1..5831
gene	/organism="Cloning vector pEX18cm" /db_xref="taxon:7315" complement(209..1630) /gene="sacB" complement(209..1630) /gene="sacB"
CDS	/codon_start=1 /transl_table=1 /product="levansucrase precursor" /protein_id="AAC27023.1" /db_xref="GI:2961141" /translation="MNIRKFAKQATVLFPTTALLAGATQAFKATNOKPKYKETYGISHTHSDMLQIPEOQNKERYQVEPDSSTIKHISAKGLDVPLOMADGTVANVHGHVHIFALAGDPKNADDTISVMEYQKVGFTSIDSKNNGRVRKDSKDPANSTIKDOTI OERGSKATFTSDKIRLFLPTDFSGKHGKQTLTTRQNVKVSADSLSINQVEDYGEESIDGDKRTIONVQQTIDEGNTSSGDHNTLDPHYVADGKHLYVEFANTGETEDGYGEESLFNVAAYGKSTSFROESQKLLQSDKRTAELANGALMEIENDYTLTKYMKPLIASNTVDEIERANVFERNKGMVLFETDSRGSKMTIDGITSNDIYMLGVSNLSGPYKPLNKTGVLKMDLDPNDVPTFTYSHFAVPAKGNVITSVYMTNRGFAKDSPTAPSEFLNLKGGKTSVKSIDSTIEQCOLTVNK" 2104..2928 /note="ori"
rep_origin	/complement(2980..3006) /note="T2"
terminator	/complement(3139..3182) /note="T1"
terminator	/complement(3186..3305) /product="Escherichia coli 5S ribosomal RNA"
rRNA	/gene="lactalpha" complement(3341..3625) /gene="lactalpha"
gene	/codon_start=1 /transl_table=1 /product="lac alpha peptide" /protein_id="AAC27024.1" /db_xref="GI:2961142"
CDS	/translation="MTMTNSSVPGDPLESTCRHASTALAVLGRDWNPGVTQLNRLAHPFASWRNSEEARTRSOQLRLSLNGEWPAWLFWRMRDEDFOPBTD" complement(4864..5397) /gene="aacC1" complement(4864..5397) /gene="aacC1"
gene	/codon_start=1 /transl_table=1 /product="gentamicin acetyl transferase" /protein_id="AAC27025.1" /db_xref="GI:2961143"
CDS	/translation="MIRSSNDVVOGSGSRPKTLGGSSMGITRTGRLPDQVKSRRALDIFREREGDVATYSGHQPPSDYLGNLRSKTFILALAFDQZAVVAGALAAVLPRFEQKRSLEYITDLAVSGEHRROGIALTALINLKHENALGAVIYVQADYGDPAVALYTKIKGIREVNHFDIDPSTAP"
BASE COUNT	1391 a 1399 c 1459 g 1582 t
ORIGIN	

QY	1	ATGACAGCGGGGAGCTCAAGAGAGACTATGGTTTGGCCATATATACAGCGGACATG	60
Db	1537	ACGAAACCAAAAGCCATATAGGAAACATAGGCACTTCCCATATATACAGCCATGATATG	1478
QY	61	CTAAATAATTCGAGGACAAACAAACAGTCCTCAATTTAAAGTGGCTCAATTCATATGATCA	120
Db	1477	CTGCAAAATCCCTGAACAGCAAAAAATGAAAAATATCAAGTTCTTGAGTTTGATTCGTC	1418
QY	121	GCAATCAAAAACATTTGATTCGGCGAAAAGGTATGATATAGTCAGGCAACTTAATAGATTTA	180
Db	1417	ACAATTTAAAAAATATCTCTTTCGCAAAAAG-----CCTG 1385	
QY	181	GATGATGGGATGCTAGCTGGCCACCTGCACAAAGCCTATGATGTACTTCGGCAAAATATATGGA	240
Db	1384	GACGTTTGCGACACTGCTGCCCTTTACAAAGCCTACGCGACTGTCCGAACATATACAGCC	1325
QY	241	TATCACATCGTCTCCGCTTTAGCAGGTGACCCCAAAAAACAGTGATGTACTCCACTTCAT	300
Db	1324	TACCAACATCGTCTTTGATTTAGCCGGAGATCTTAAAAATGGCGATGACACTTCGATTTAC	1265
QY	301	TTATTCATCAAAAAGTCGGTGATATACATTCGATACAGCTGGAAAAATTCGCGAAGGTA	360
Db	1264	ATGTTCTATCAAAAAGTCGGCAAACTTATTTACAGCTGGA AAAACGCTGCGCGCTC	1205
QY	361	TTTGAAGATATGAGTAATATTTGTCCCAAAATGATCCGATCTTAAATATCAAAACAGAG	420
Db	1204	TTTAAAGACAGCAAAATTCGATGCAAAATGATTTCTATCTAAAAAGCAACACAGAA	1145
QY	421	TGTCAGGTTGTGCTACTTTTACCAAAGATGGCCAAATGCGTATTATCTATACAGATTAC	480
Db	1144	TGGTCAGGTTGAGCCACTTTTACATCTGACGGA AAAATCCGTTTATCTTACAGGATTTTC	1085
QY	481	TCAGGTAATCCCTGAAGATGGTGAAACCGGTCTGTTAAACCAAAATCTTCAACTGCTCAA	540
Db	1084	TCGCGTAAACCTTA-----CGGCAACGAACACATGACACATGCAACA 1043	
QY	541	GTAACCTTATCCACGCGGATGCACTACCTTAAAGTCATGGAATTCATGATCATAA	600
Db	1042	GTTAACGATATGAGCATGACAGACGCTCT---TTGAACATCAACGGGTGATGAGAGATTATAAA	985
QY	601	TCTGCTCTTGTGAGCGGAGGATACAGTTTATCAAAATATTCAGCAATTTATGATGAA	660
Db	985	TCAATCTTTGA---CGTGACGGA AAAACGATATCAAAATGTACAGCACTTATGATGAA	929
QY	661	GGCAGTGAGATTGAGTGATTAACCATACTTTAAGAGACCTCACTATGTTGAAGATAG	720
Db	928	GGCAACTACACTGACGGCAACACCATATCGTGAGAGATCTCTCACTAGTGAAGATAAA	869
QY	721	GGCATTAATATCTCTGCTTTGAAGCGCAATTCGTGAACAACAGATGGTTATCAAGCGAT	780
Db	868	GGCCACAATATCTAGTATTTTGAAGCAAAACCTGGAATGAGATGGCTACCAAGGCA	809
QY	781	CAGTCTTCAATTAATAAGCTTACTATGAGCGGAGAGACGCTCTCTTCCAGAAATGAAAA	840
Db	808	GAATCTTATTTAACAAGCATACTATGCGCAAAAGACATATCTTCCGTCAGAAAGT	749
QY	841	AATTAAGTGCCTTCAAAGCTCTAAAAAACAATTCCTCTTTAGGGAATGGTGCAATAGGC	900
Db	748	CAAAAACCTTCGCAAAAGGATAAAAAGCGCAGGCTGAGTTAGCAAAACGGCGCTCTCGGT	689
QY	901	ATTGTTGAATTTGGCGGATGACTATACAGTGAAGAGTGTATGAACCATTTAGTGCAATCA	960
Db	628	AACACAGTACAGATGAATTTGAACGCGCGCAACCTTTTAAATGAACGCGCAATGCTAT	569
QY	961	AACACAGTAGCAGATGAAGTGAAGCGCGCAATATATTTAAATGAATTAATTAATGGTAT	1020
Db	1021	CTATTCACGATTCACAGAGTCCAAAATGACGAGTATGATGAATTAACGACAAAGATGT	1080
QY	568	CTGTTACGACTCTCCCGGATCAAAAATGACATGACATGACGCAATTAAGCTTAACGATAT	509



QY	1081	TATATGCTAGAGCCCGGAGCGGACTCCCTTAATGTGGCCACACAAACCGGATTAATGAACCT	1140		
Db	508	TACATGCTGGTTATGTGTTTCTTAATCTTTTACTGAGCCCATCAAGCGCTGAACAAACT	449		
QY	1141	GGACTTGTATTTGAACATGATCTTGACCCCTGCTATCTCAACACACACTTACTCTCATTTGC	1200		
Db	448	GGCCTTGTTAAATAATGATCTTGATCTTACAGATGTAACCTTTACTTACTCACTTC	389		
QY	1201	GGTATCCCGCACCTGAAAGGTAATATATGTGTTACTCATCAAGTTATATGACAAATAGAGCC	1260		
Db	388	GCTGTACCTCAAGGAAAGGAAACAAGTCGTGATTACAAGCTATATGACAAACAGAGA	329		
QY	1261	TTTCATCCAGAACATCATCTCTCACCCTGCGGACACAGCTTGGGGTTATATTAAGGCTCT	1320		
Db	328	TTCTACGACGACCAACATCAACGTTTGCGGCTTAGCTTCTGTGACATCAACGAAG	269		
QY	1321	GACACATCTGGAGGAGAAATAG	1343		
Db	268	AAACATCTGTGTCAAAAGACAG	246		
RESULT 12	AF004910/c	5842 bp	DNA	Linear	SYN 01-OCT-1999
LOCUS	AF004910				
DEFINITION	Cloning vector pEX18T, complete sequence.				
ACCESSION	AF004910				
VERSION	AF004910.1	GI:2218142			
KEYWORDS	Cloning vector pEX18T.				
SOURCE	Cloning vector pEX18T				
ORGANISM	artificial sequences; vectors.				
REFERENCE	1 (bases 1 to 5842)				
AUTHORS	Hoang,T.T. and Schweizer,H.P.				
TITLE	Improved vectors and mobilizer strains for gene replacement in				
JOURNAL	Pseudomonas aeruginosa				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 5842)				
TITLE	Hoang,T.T. and Schweizer,H.P.				
JOURNAL	Direct Submission				
FEATURES	Submitted (20-MAY-1997) Microbiology, Colorado State University,				
SOURCE	Microbiology Building B16, Fort Collins, CO 80523, USA				
	Location/Qualifiers				
	1. 5842				
	/organism="Cloning vector pEX18T"				
	/db_xref="taxon:61449"				
	209. 1630				
	/gene="sacB"				
	complement(209. 1630)				
	/codon_start=1				
	/transl_table=1				
	/product="levansucrase"				
	/protein_id="AAB61667.1"				
	/db_xref="GI:2218143"				
	translation="MNKKRKAKQATVLTFTTALLAGATQAFATENOQPKPKETYGIS				
	HTRHDLQIPEBOQNEKQVPEPDSSTINISAKGLDWDSPLONDAGTVANHYG				
	YHVFALADPKNADDTISYFYOKVGETSIDSWKNAGRVFKDSKDPDANDSLIKDOT				
	QEMSGSATPTSDGKILRLFYDFSGKHGKOTLTTAQVNASDSLSINQEDYKSI				
	DGCKRTYQNFQIDEGNYSAGDHHTLRDHYVEDKSHKYLVEFANVTGTEDEYGEBS				
	LFNKAYYGSTSPROESOKLDSKRTAEIANGALGMEIENDYTLKKVMKPLNS				
	MYTALDERANVFRMNGKRYLFTDSDRSKSKTIDSTISNDIYMLCYSNSLTPKPKPLN				
	KTGLVLAAMDLPDNDVTTYTFSHFAVPAOKAGNNVITSYTMNGFYADKOSTFAPSFLLN				
	IKCKKTSVVDKSDIIEOQLTVNK"				
	2104. 2928				
	/note="region containing oriT"				
	complement(2980. 3007)				
	/note="T2"				
	complement(3014. 3139)				
	/note="T1"				
	complement(3018. 3305)				
	/product="5S ribosomal RNA"				
	/note="from Escherichia coli"				
	3341. 3625				
	/gene="lacZalpha"				

CDS		complement(334..1..3625)
		/codon_start=1
		/transl_table=1
		/product="Iac alpha peptide"
		/protein_id="AAB61668.1"
		/db_xref="GI:2218144"
		/translation="MTMTNTSSSPGDPLESTCRHSALAVLVQDRDMENPGYTQLN RLAHPPPAWNRNSEARTDKPSOOLSLNGEMRWLFWRRDFQPTD"
gene		4782..5642
		/gene="bla"
		complement(4782..5642)
		/codon_start=1
		/transl_table=1
		/product="beta-lactamase"
		/protein_id="AAB61669.1"
		/db_xref="GI:2218145"
		/translation="MSIOHFRVALIPPEAFCLPVFAHPETLVKVAEDQLGARVYG IELDNISKEIFSRPREFRPFMMSTFVKYLGGAVLSLRDAEOELGRIRHSONDIYF YSPTEKRHLTDGMTVRELCSAITIMSDTANLLTTIGSEKLTAFLHNHGDTVR DRWEPLNEALIPNDERDTPVMATYLTAKLLTGTELTLASRODLIDMEADKVAIGPI LRSLAPGMWFIADKSFGAGERSRGIIIALGDGPISRIVLYITTSQTMDERNROIIE EGSLIKIHW"
BASE COUNT	1400 a	1392 c 1435 g 1615 t
ORIGIN		
Query Match	42.4%	Score 578.6; DB 12; Length 5642;
Best Local Similarity	67.2%;	Pred. No. 1.5e-123;
Matches 903;	Conservative 0;	Mismatches 389; Indels 51; Gaps 4;
Oy	1	ATGAACACGGGGGACTACAAGSAGACATATGTTTGGCCATATTACACGGCGTCACATG 60
Dd	1537	ACGAACCAAAAGCCATTATAGGAACCAATACGCCATTTCTCCCATATTAACGCACATATG 147B
Oy	61	CTAAAATAATCCAGGACACAAACACAGTCCCTCAATTTAAAGGCCCAATCAATCATCA 120
Dd	1477	CTGCAATACCCTGGAACACACAAAAAATATAAAATATCAAGTTCCGAGATTGATTCGTCC 141B
Oy	121	GCAATCAAAACATTTGATTCGGCAAAGGGATATGATTAAGTCAGGCAACTAATAGATTTA 180
Dd	1417	ACAATTAATAAATATCTCTTGTGCAAAAG-----CCWG 138B
Oy	181	GATGATGGGATAGCTGGCCACATGCAAAACGCTGATGGTACTGCGGCAATTTATCATGA 240
Dd	1384	GAGCTTTGGGACAGTGGCCATTACAAACGCTGACGGCAGCTGTCGCAAACTATCACGGC 132B
Oy	241	TATCATCATGCTCCGCTTTAGCAGGTATGCCAANAACAGTAGATGATGCTCACTTGAT 300
Dd	1334	TACCACTATCGTCTTTGCAATTAGCCGGAATCCTTA AAAATGCGGATGACACATCGATTAC 1265B
Oy	301	TTATTTATCAAAAAGTGGGTATACATGATTTGACAGCTGGAAAAATGCTGAAGACTA 360
Dd	1264	ATGTTCTATCAAAAAGTGGGAAACCTCTATTGACACTGGAAAAAGCGTGGCGCGCTC 1205B
Oy	361	TTTGAGATATGAGATAAATTTTGTTCCAATGATCGGTATCTTAAATATCAAACACAGAG 420
Dd	1204	TTTAAAGACAGCGCAAAATTTGATGCAATATATCTTCAAAAAGACCAACACAAAGAA 1145B
Oy	421	TGGTCAGGTTCTGACTTTTAACCAAGAATGGCCAAAGCCGTTTATCTATACAGATTAC 480
Dd	1144	TGGTCAGGTTTACGACATTTACATCTGACGGAAAAATCCGTTTATCTACACTGATTTC 1085B
Oy	481	TCAGTATCCGTAAGATGATGGAACCGTGTGCTGTAACCAAAATCTATTCAACTCTCA 540
Dd	1084	TCGGTAAACATTA-----CGGCANAACAACCTGACAACTGCACAA 104
Oy	541	GTAACCTTATCCCACGCGGATGACACTATTAAGTCGATGAGATCTGATCATATA 600
Dd	1042	GTTAAACGATATCAGATCAGACAGCTCT---TTGAACATCAACCGGTGAGAGATTATATA 986
Oy	601	TCTGCTTTGANTGGGAGAGGATCACTTATATCAAAATATTCAGCAATTTATGATGATA 660
Dd	985	TCAACTTTTA---CGGTACGGAAAAACGATACAAATATGACACGCACTTCATCATGATA 929

```

OY 661 GGCAAGTGGATTTTCAGTGTATACCATCTTTAAGAGACCCCTACTATGTGAAGATAG 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 928 GGCACTACAGTACGGGACACACATACCGTGCAGATCCCTACCTAGTAGAATATA 869
OY 721 GGCCTAATATCTTGTCTTTGAGCGAATCTGCAACACAGATGTTATCAGGCCAT 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 868 GGCCTAATATCTGATTTGAGCAACACACTGCAAGATGCTACACAGGCCAA 809
OY 781 CAGTCTTCATTAATTAAGCTTACTATGCGGCAATGACGCTCTTCCAGATGAAAA 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 808 GATCTTTATTTAACAAGCATACTATGCAAAACACATCTCTTCCGTCAGAAAGT 749
OY 841 AATAAAGCTCTCAAGTCTTAACAAAACAAATTCCTTTAGCGAATGTGCATTAGGC 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 748 CAAAAAGCTTGCAGAAAGGATATAAAAGCAACGCGTGAGTTAGCAAGGGCGCTCGGT 689
OY 901 ATTGTGAATTTGGCGGATGACTATACATGAAAGTGTATGAACCATTAAGCATCA 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 688 ATGATTTGAGCTAAAGATGATTAACACTGAAAAAGATGAAAAAGCGCTGATTCATCT 629
OY 961 AACACAGTAGACGATGAGTGAAGCGCCCAATATATATTAATGAATTAATGATAT 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 628 AACACAGTAGACGATGAGTGAAGCGCCCAATATATTAATGAATTAATGATAT 569
OY 1021 CTATTACAGGATTCAGAGATCCAAATGACGAGTGAATTAACGACAAAGATGT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 568 CTGTCTACTGATCCCGCGGATCAAAATGACGATTAAGCGGATACGCTAACGATAT 509
OY 1081 TATATGAGAGGCGCGGAGCGGATCTTAATGCGCCACACACACCCGATTAATGAAT 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 508 TACACTGTTGTTATGTTCTTAATCTTAACTGCCCCATACACCCGCTGAACAAACT 449
OY 1141 GGACTTGTATTAACATGATCTTGACCCCTGCTGATCTCACACACTTACTCTCATTC 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 448 GGCCTTGTGTTAAATGATGATCTTATCTTAACGATGTAACCTTTACTACACACTTC 389
OY 1201 GGTATCCGCAACCTGGAAGTAAATGTTGTTACTACAGATTTATAGCAATAGAGGC 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 388 GCTGTACTCTCAAGCGCAAGCAACATGTCTGATTACAGCTATATGACAAACAGAGA 329
OY 1261 TTTATTCAGAAACATCACTCTACCTGCGGAGCAAGCTTGGGTTAATTAAGCGTCT 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 329 TTTACGAGCAACAAACATCAACGTTTGGCCCTAGCTTCTGCTGACATCAAAAGCAAG 269
OY 1321 GACACATCTGGAGAGAAATAG 1343
DB 268 AAACATCTGTTGCAAGACAG 246

RESULT 13
XXU17500/c 5846 bp DNA circular SYN 18-DEC-1994
LOCUS Cloning vector pEX100T, complete sequence.
DEFINITION U17500
ACCESSION U17500.1 GI:603175
VERSION beta-lactamase (bla); lac alpha peptide (lacZalpha); levanucrase
KEYWORDS (sacB); origin of transfer (oriT); transcription terminators;
SOURCE megaluclease I-SceI.
ORGANISM Cloning vector pEX100T.
REFERENCE 1 (bases 1 to 5846)
AUTHORS Schweizer,H.P. and Hoang,T.T.
TITLE An improved system for gene replacement and xyle fusion analysis in
Pseudomonas aeruginosa
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5846)
AUTHORS Schweizer,H.P.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-1994) Herbert P. Schweizer, Microbiology &
Infectious Diseases, University of Calgary, 3330 Hospital Drive
N.W., Calgary, Alberta T2N 4N1, Canada
FEATURES Location/Qualifiers

```

```

source
1..5846
/organism="Cloning vector pEX100T"
/species="Escherichia coli"
/db_xref="taxon:37970"
/plasmid="pEX100T"
/translation="209..1630"
/compelment(209..1630)
/gene="sacB"
/compelment(209..1630)
/gene="sacB"
/compelment(209..1630)
/gene="sacB"
/codon_start=1
/transl_table=11
/product="levanucrase"
/protein_id="AA57325.1"
/db_xref="GI:603178"
/translation="MNIRKFAKQNTVLFETALLAGATQAFAKETNOKPKYETGIS
HTRHDLQIPKQKNEKQYPERDSSRTIKSSAKGLDWSPLONADGPTVANYG
YHIVFALAGDPRKADDTSTHYFYKQVETSIDSKNNKGRYRKDSKDDANDILKDOT
QEMSGATFTSDGK1RLFYDTSGRKHGKOTLTAQVNSASDSSLININVEDYKSLF
DGDRTYQNVQOFLDEGNYSGDHHTLRDPHYVDKHKYLVFPAANTGDEGYOGEES
LFNKAYYKSTSPFROESOKLLOSDKRTAEALGALGMIELANDYTLKKVMPPLIAS
NTVIDERAVFMKNGKMYLFTOSRGSKMTIDGITSNDIYMLGYVNSLTPGPKPLN
KTGLVTKMDLPNDPVTFTYSHFAYPQAKGNVITISYMTNRFYADKOSTFAPSFLN
IKKRTSYVRKDSILEGOLTVNK"
2104..2928
/region="region containing oriT"
/misc_feature
/compelment(2984..3309)
/feature="E. coli 5S RNA and terminators"
/compelment(2984..3011)
/feature="r2"
/compelment(3018..3143)
/feature="r1"
/compelment(3022..3309)
/product="5S ribosomal RNA"
/compelment(3345..3629)
/gene="lactalpha"
/compelment(3345..3629)
/gene="lactalpha"
/codon_start=1
/transl_table=11
/product="lac alpha peptide"
/protein_id="AA57324.1"
/db_xref="GI:603177"
/translation="MTMTTPSSRDNRVYIPGRDNRYMSLALAVLQRDWENPGVTOLN
RLAHPFASWRNSEARTRPSOQLRSLNENRAMPFWRMEDFQPDPTD"
3565..3605
/feature="polylinker"
/compelment(4786..5646)
/gene="bla"
/compelment(4786..5646)
/gene="bla"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AA57323.1"
/db_xref="GI:603176"
/translation="MSIQHFRVALIPFFAFLPFAHPETLVYVKQAEDQOLGARGVY
IELDLSNGKILIESFRPEERPMSTFVLLCGAVLSIDGQEOGLGRIRIHYSQNDLVE
YSPVTEKHLTDGMYRELCSAATMSNTAANLLTTIGCKELTAELAHNNGDVTRL
DRWPELNEATPNDERTPTVMATTLRLKLTGELLTLASROOLIMDEADKVAQPL
LRSLPAGMTADPSGAGERGSRGIIAALGDPGKPSRIYVITYTGGSATMDERRQQLA
ETGASLIRHM"

BASE COUNT 1400 a 1395 c 1430 g 1621 t
ORIGIN
Query Match 42.4% Score 578.6 DB 12 Length 5846;
Best Local Similarity 67.2% Pred. No. 1.5e-123;
Matches 903; Conservative 0; Mismatches 389; Indels 51; Gaps 4;
OY 1 ATGAACAGCGGGGACTACAGCAAGACATGTTTCCCATATTTACACGCGCTGACATG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1537 ACGAACCAAAAGCCATATAGAAACATACGCGCATTTCCATATTACGCGCATGATATG 1478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 61 CTAAAAATTCAGGACAAACAAAGCTCTCAATTTAAAGTGCTCAATTCATGATCA 120

```

```

Db      1477 CTGCAAACTCCCTGGAAGCAAAAAAATGAAAAATATCAAGTTCCGTGAGTTGATTCCTGCC
Oy      121  GCATTAACAAAACATTTGATGGCAAAAGGGATATGATAGTCAGGCAACTTAATGATTTA 180
Db      1417 ACAATTAATAAATATCTCTTCTGCAAAAG-----CCTG 1385
Oy      181  GATGTATGGATAGTACCTGGCCACTGCAAAACCGTATGATGCTGGCAAAATATCATGGA 240
Db      1384  GACGTTGGGACAGCTGGCCATTAACAAAGCTGACGGCACTGTCGCAAACTATCACGGC 1325
Oy      241  TATCAATCGTCTCCGCTTTAGCAGGTGACCAAAAAACAGTATGATATCCACTTCAT 300
Db      1324  TACCACATCGTCTTGGCATTTAGCCGGAGATCTTAAATAATGGGATGACATCATGTTTAC 1265
Oy      301  TTATTTCTAACAAGATGCGATATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db      1264  ATGTTCTATCAAAAAGTGGCAAACTTCTATGACAGCTGGAATAACGCTGGCCGCTC 1205
Oy      361  TTGAGATATGATGATTAATTTGTTCCAAATGATCCGATCTTAAATATCAACACAGAG 420
Db      1204  TTTAAGACAGCAGCAAAATTTGATGCAAAATGATTTCTAATAAAGACCAACACAGAA 1145
Oy      421  TGGTCAGGTTCTGCTACTTTAACCAAGATGCCAAGTCCGTTATTTCTATACAGATTAC 480
Db      1144  TGGTCAGGTTCTGACACATTTTACATCTGACGAAAAATCCGTTATTTCTACATGATTTC 1085
Oy      481  TCAGTATCTCTGAAGATGATGGAACCGGTGCTGATCAAAATCATTTCAAGCTCTCA 540
Db      1084  TCCGTTAAACATTA-----CGGCAACAAACACTGACAACTGACACAA 1043
Oy      541  GTAACCTATATCCAGCCGATGACATACCTTAAGTCAGTGAAGTATGATCATATA 600
Db      1042  GTTAACGATATGAGATGACAGACGCTCT--TTGAACATCAACAGGTATGAGATATA 986
Oy      601  TCTGCTTTGATGGCGGACGATGACGTTATCAAAATATTCAGCAATTTATGATGAA 660
Db      985  TCAATCTTTGA---CGGTGACGGAATAACGTAATAATGTAACAGCACTTATGATGAA 929
Oy      661  GGCAAGTGAATTTAGGATGATTAACCATCTTAAGAGACCTCTATGTAAGATTAAG 720
Db      928  GGCAACATACAGCTAGGCGACACACATACGCTGAGAGATCTCTACATGATGATATA 869
Oy      721  GGCAATAATATCTTCTCTTGAAGCAATATGTAACACAGATGTTATCAAGCGCAT 780
Db      868  GGCCACAATATCTAGTATTTGAAGCAAACTGGAATGGAATGGCTACCAAGCGAA 809
Oy      781  CAGTCTTTCAATATAAGCTTACTATGCGGGAAGTACGCTCTTCTTCCACATGAATA 840
Db      808  GAATCTTTATTTAAACAAAGCACTATGCAAAAGCACATCATCTTCCGTCAGAAAGT 749
Oy      841  AATAACGCTTCTTAAAGCTTAAACAAACAAATTTGCTTTAAGGAATGGTCAATTAGC 900
Db      748  CAAAACTCTTCTGCAAAAGCGATATAAAGCGACGCTGATGATGACAAAGCGCTCTCGT 689
Oy      901  ATTTGTAATTTGGCCGATGACTATACAGTGAAGATGTTAATAACCATATGATGCAATCA 960
Db      688  ATGATTTAGATAAAGATGATTAACACTGAAAAAGTGAAGTGAACCGCTGATTGCACT 629
Oy      961  AACACAGTACAGATGATGATGGAAGCGGCAATATATTTAATAATGAATTAATGAT 1020
Db      628  AACACAGTACAGATGATGATGGAAGCGGCAATCTTTAAATGAAGCGCAATGAT 569
Oy      1021  CTATTCAGGATTTAAGAGATGCAAAATGAGATGATGATTAACAGCAAGATTT 1080
Db      568  CTGTTCACTGATCTCCCGGATCAAAAATGACGATGACGCAATTAAGTCTTAACGATAT 509
Oy      1081  TATATGCTAGGCGCGGAGGAGCTCTTAATATGCGCACAAACCGGATTAATGAAT 1140
Db      508  TACATGCTGTTGATGTTTCTAATTTCTTAATCTGACCATCAAGCGCTGACAAACAT 449
Oy      1141  GGAATTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200

```

```

Db      448  GGCCCTTGTTAAATAATGATCTTGATCTTAACGATGATTAACCTTACTTACTCACACTTC 389
Oy      1201  GGTATCCCGCACCCGAGGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
Db      388  GCTGATCTTCAACGCAAAAGGAAACATGCTGATTAACACTTATATGACAAACAGAGA 329
Oy      1261  TTTATCCAGAACATCACTTACCTGCGGAGCAAGCTTGGGCTTAATATTAAGGGCTCT 1320
Db      328  TTTATCCAGAACAAACATCAACGTTTCCGCTGCTTCCGTCGATCAATCAAAAGCGCAG 269
Oy      1321  GACACATCTGAGAGGAAATAATAG 1343
Db      268  AAACATCTGTTGTCAAAGACAG 246

RESULT 14
SYNPRIC 5941 bp DNA circular SYN 27-APR-1993
LOCUS Cloning vector (PRL278) for sacB-mediated positive selection for
DEFINITION double recombinants in gram-negative bacteria.
ACCESSION L05083
VERSION L05083.1 GI:209132
KEYWORDS Cloning vector.
SOURCE Cloning vector DNA.
ORGANISM unidentified cloning vector
REFERENCE 1 (bases 1 to 5941)
AUTHORS Cai, Y.
TITLE Molecular genetic approaches towards the understanding of
heterocyst differentiation and pattern formation in the
cyanobacterium Anabaena sp
unpublished (1992)

JOURNAL Black, T.A., Cai, Y. and Wolk, C.P.
REFERENCE Spatial expression and autoregulation of hetr, a gene involved in
AUTHORS the control of heterocyst development in Anabaena
TITLE unpublished (1992)
FEATURES
source Location/Qualifiers
1..5941
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
complement(473..1894)
/feature="sacB from B. subtilis; ORF"
/codon_start=1
/transl_table=11
/protein_id="AA172307.1"
/db_xref="GI:209133"
/transl_table=11
/transl_table="MNIRKFAKQATVITFTTALLAGATQAFAKETNOKPYKETYGIS
HITRHMLOIPEOQKNEKYQVPEPDSSTIKNISAKGLEWDSMPLOADGTVANHG
YHIVFALAGDKNDKNTSIVMFYKVGSTSIDSKNAGRVKDSKFDANDSILKDOT
QEMSGSATFTSDGKIRLEYPDSEKHKHOKLTTAAYVNVASDSSINNGVEDYKSI
DGDCKTQONNOOFTDEGNSSGDNHITLDPHYVEDKGIKIVFPANNGTEGYGEB
LFPKAYGKSTSPFROESOKLLOSDKRTAELANGALMELNDYTLKVKMPLAS
NTVTEIERANVAFKMGKMWLEFDSRSKMTIDITSNDIYMLGVYSLSLGPYKPLN
IKGKTSYVKDLSILEOGOLTVNK"
2749..3543
/feature="kanamycin and neomycin resistance determinant
(unp1)"
/codon_start=1
/transl_table=11
/protein_id="AA172308.1"
/db_xref="GI:209134"
/transl_table="MIEDGILHAGSPAMVERLFGYMAOQTIGCSDAAVPRLSAQR
PVLYVKTDLSALNELDDEARLSMLATTPCAAVLDVYEAQRDMILLGEVGDLE
LSHLDAPEKVISINADHARRLTLDPACTPDDAKKHIERARMEGLVDODDDE
EHQGLAPELFAELKARMPDDEDLVTHGDACTENINVENRFGFTIDCGRLVADRY
ODIALATRDIEELGEMADRFVLVYGIAPDSORIAFYRLDDEF"
3558..4815
/feature="multiple cloning sites"
/misc_feature misc_feature
4816..5941
/feature="oriV and oriT genes from Plasmid pMB1"
BASE COUNT 1419 a 1329 c 1471 g 1722 t

```

## ORIGIN

Query Match 42.4% Score 578.6; DB 12; Length 5941;

Best Local Similarity 67.2% Pred. No. 1.5e-123; Mismatches 389; Indels 51; Gaps 4;

Matches 903; Conservative 0; Mismatches 389; Indels 51; Gaps 4;

```

OY 1 ATGAACAGCGGGGACACAGAGAGACTGTTTGGCCATATTATACAGCGCGTACATG 60
DB 1801 ACGAACAAGGACATATAGAGAAACATACGCGCATTTCCATATTACACGCGCATGATG 1742
OY 61 CTAATAATCCAGGACACAAACAGTCCCTCAATTAAAGTCCCAATTCATGATCA 120
DB 1741 CTGCAATCCCTGGAACGCAAAAAAATAATCAAGTTCCGTAATTCGATTCGTC 1682
OY 121 GCAATCAAAAACATTTGATTCGGCAAAAAGGATGATAGTACAGCACTTAATAGATT 180
DB 1681 ACANTTAATAATATCTCTTCGCAAAAG-----CCTG 1649
OY 181 GATGTAGGATAGCTGGCCACTGCAAAAAGCTGATGTTACTGGCGCAATTTATCAG 240
DB 1648 GAGGTTTGGACAGCTGGCCACTTACAAAACGCTGCGCACTGTCGCAACATATCAG 1589
OY 241 TATACATCGTCCGCTTTAGCGAGTGACCCAAAACAGTATGATACCTCCACTTAT 300
DB 1588 TACACATCGTCTTTGATTAAGCCGAGATCCTTAATAATGCGATGACATCGATTAC 1529
OY 301 TTAATCTATCAAAAAGTGGTGAATACATGATTCACAGCTGGAAAAATGCTGAAGATA 360
DB 1528 ATGTTCTATCAAAAAGTGGGGAACCTCTATGACACCTGGAAGAAAGCTGGCGGCTC 1469
OY 361 TTTGAGATATGATTAATTTGTTCCAAATGATCCGTATCTTAATATCAACACAGAG 420
DB 1468 TTTAAAGACAGCGCAAAATTCGATGCAATATCTTATCAAAAGACCAACACMAAA 1409
OY 421 TGTGAGGTGCTGCTACTTTAAACCAAGTGGCCAAAGCTTTTATATACAGATTAC 480
DB 1408 TGGTACAGTTACGACCATTTACATCTGACGAAAAATCCGTTTATCTACAGATTTC 1349
OY 481 TCAGTAACTCTGAAAGATGTGGAACCGTCTGTGAACCAATCATTTCAACTGCTCAA 540
DB 1348 TCCGTAACATTA-----CGGCAAAACAAACACACACACTGACACAA 1307
OY 541 GTAACCTATCCCAAGCCGATGACATCTTAAGTCGATGAGTATCTGATATATA 600
DB 1306 GTTAACGATACGATCAGACAGCTCT--TTGAACATCAACGCGTGAAGATTATAA 1250
OY 601 TCTGCTTTGATGGCGAGACGCTACACTTATCAAAATATTCACCAATTTATCATGAA 660
DB 1249 TCATCTTTGA---CGGTGACGGAACAAACGATATCAAAATGTACAGCACTTCATG 1193
OY 661 GGCAGTGTGATTCAGTGTATACCATATCTTAAGAGACCTCACTATGTTGAAGATA 720
DB 1192 GGCAGTGTGATTCAGTGTATACCATATCTTAAGAGACCTCACTATGTTGAAGATA 1133
OY 721 GGCATTAATATCTGCTTTGGAAGCAATCTGGAACACAGATGTTATCAGGCGAT 780
DB 1132 GGCACCAAAATCTAGTATTTGAAGCAACACTGGAAGTGAAGAGGCTACCAAGCGCAA 1073
OY 781 CAGTCTTCATTAATTAAGCTTACTATGCGGAGTGAAGTCTCTCTTCAGAAATGA 840
DB 1072 GAATCTTTATTAACAAAGCATACTATGCGCAAAAGCATATCTTTCCGTCAGAAAGT 1013
OY 841 AATAAAGCTGCTTCAAGTCTTAACAAACAAATTTGCTTTAGCAAGATGTCATTAAG 900
DB 1012 CAATAAAGCTTCAAGGATTAACAAAGCAAGCGGTGAGTTAGCAAGGCGGCTCGCT 953
OY 901 ATTGTTGATTTGGCGGATGATATACAGTGAAGATGTTATGAACCAATATGTCATCA 960
DB 952 ATGATTTGAGCTTAACGATGATTAACACTGAAGAAAGTGAAGAAACCCGATGATCAT 893
OY 961 AACACAGATGATGAAGTGAAGCGGCGCAATATATTAATAAGATTAATATGAT 1020
DB 892 AACACAGTATGATGAAGTGAAGCGGCGCAATCTTTTAATAAGAACGCAAGATGAT 833

```

```

OY 1021 CTATTCAGGATTCAGAGATTCACAAATGACAGATGATGATTAACGACAAAGATGTT 1080
DB 832 CTGTTCACTGACTCCCGGGATCAAAATATGACATGTAGCGCATTAAGTCTAAGATATT 773
OY 1081 TATATGCTAGGGCCCGGAGCGACTCTTAATATGCGCACAAACCCGATTAATGAAT 1140
DB 772 TACATGCTGTTGATTTGTTTCAATTTCTTAATCTGAGCCCATATCAAGCCGTCGA 713
OY 1141 GGACTGTGATGAACATGATCTGACCCCTGATATGACATGACACACCTTACTCTCATTC 1200
DB 712 GGCCTGTGTTAAATGATCTGATCTAAGATGTAACCTTACTTACTTACTTACTTACT 653
OY 1201 GGTATCCCGCACCCCTGGAATTAATGTTGTTACTACACAGTTATATGACGATAGAG 1260
DB 652 GCTGACTCAAGGAAGAAAGAAACATATGCTGATTAACACTTATATGACAAACAGAGA 593
OY 1261 TTTATCCAGAACATCACTCTACCTGCGGACAAAGCTTGGGGTTAATATTAAGGCT 1320
DB 592 TTTACGCGAGCAAAACATCAACGTTTGGCCAGACTTCTGCTGACATCAAAAGCAG 533
OY 1321 GACACATCTGAGAGAGAAATAG 1343
DB 532 AAAACATCTGTTGTCAAAGACAG 510

RESULT 15
SYNPLA/C 6347 bp DNA circular SYN 27-APR-1993
DEFINITION Cloning vector (pRL271) for sacB-mediated positive selection for double recombinants in gram-negative bacteria.
VERSION 105081
KEYWORDS L05081.1 GI:209125
SOURCE cloning vector DNA.
ORGANISM unidentified cloning vector.
REFERENCE 1 (bases 1 to 6347)
AUTHORS Cal, Y.
TITLE Molecular genetic approaches towards the understanding of heterocyst differentiation and pattern formation in the cyanobacterium Anabaena sp
JOURNAL unpublished (1992)
REFERENCE 2 (sites)
AUTHORS Black, T.A., Cal, Y. and Wolk, C.P.
TITLE Spatial expression and autoregulation of heter, a gene involved in the control of heterocyst development in Anabaena
JOURNAL unpublished (1992)
FEATURES
    source
        1..6347
            /organism="unidentified cloning vector"
            /db_xref="taxon:45196"
            complement(473..1894)
            /note="sacB gene from B. subtilis; ORF"
    CDS
        1..6347
            /transl_table=1
            /protein_id="AA072302.1"
            /db_xref="GI:209126"
            /translation="MNIRKFAKONATVLTFTTALLAGTQAFKATNOKPKYKTYGIS
            HITHDMLOIPEQOKNEKYVPERPDSITIKKISSAKLEWDSPLNOACTVANYHC
            YHIVFALAGDKRNDSTIYWKVQKVEISIDSKNAGRFKDSKFDANDSTILADDT
            QEWGSAITFTSDGKIRLEVTDFSGKHGKQTLTAOVNVSADSSLLINVEDYKSLF
            DGDGKYOVNOVOFIDEENYSSGDHPLRDPHYVDGKHGKLYLVEPANGTEDEGYGEE
            LFNKAYYKSTSFRESOKLJSDKRTKRLAANGALGMEIENDYTLKKYMKPLIS
            NTVNDELRAVNEFMNKKWLYFDSRSGSKMTIDGTSNDIYMLGYVNSLTLGPYKPLN
            KTGVLKMDLDPNDVTFTYSFAYPQAKNNVITTSYTNRGTAFDASTPAPSFLLN
            IKGRTSVVKDSILLEGOLTVNK"
            complement(2393..3127)
            /note="erythromycin resistance determinant"
            /codon_start=1
            /transl_table=1
            /protein_id="AA072303.1"
            /db_xref="GI:209127"

```

```

/translation="NMEKNIKHSQNFITSKNIDKIMTINRLNEHNDNIFEIGSKGH
TELVKRCNEVTAIEIDHKLCKTTEKNKLVNDHNOVLNKLIDJFKPKNOXYIYENI
PTEVSTDIRKIVEDISIANEITYLVEYGAKRLNTRSLRLLALMEVDSIISLMPR
EYHPRKXNSLSIRSRKRSRISHDKOKYNTFYWKVKNKEKIKFTPNQFNLSLKH
AGIDDLNNISFEDEFLSLFNSYLKFNK"
3503.4162
/feature="chloramphenicol resistance determinant"
/codon_start=1
/transl_table=11
/protein_id="AA72304.1"
/db_xref="GI:209128"
/translation="MEKITGTTTVDISQMRKNEFEAFQSAQCTYNOTVOLDITAF
LKVKNKHKRFPALFTHILRLMNAHPEFRMAKDELIMDVSHCYTWFHBOETFE
SSLSMEYHDDRFOLHIYODVACYGENIAYPKFTEENFEYSANPWFSEFELNV
ANMDFEAFVFTMGKYITOGDKVLMPLAIQVHNHAYCDGFHVGSMLEIDQYCEMGG
A"
misc_feature 4252..5221
             /note="multiple cloning sites"
misc_feature 5222..6347
             /note="ori" and ori genes from plasmid pMB1"
BASE COUNT 1649 a 1305 c 1404 g 1989 t
ORIGIN
Query Match 42.4%; Score 578.6; DB 12; Length 6347;
Best Local Similarity 67.2%; Pred. No. 1,5e-123;
Matches 903; Conservative 0; Mismatches 389; Indels 51; Gaps 4;

QY 1 ATGACAGCGGGGACATACAGAAAGACTATGTTGGCCATTTACAGCGGTGACATG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1801 ACCAACCAAAAGCCATTAAGAAACATACGACATATTCACCCATATG 1742
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CTAAATTTCCAGACAAACAAACAGTCTCAATTTAAAGTGCCTCAATTCATGATCA 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1741 CTGCAATTCCTTGAACAGCAAAAGAAATGAAATATCAAGTTCTGATTCGATTCGTC 1682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GCAATCAAAACATTTGATTCGCAAAAGGATGATTAAGTACGACCACTTAATGATTTA 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1681 ACATTTAAAAATTTCTTCTCGCAAAAG-----CCTG 1649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GATGTATGGATGCTGGCCACTGCAAAAGCTGATGCTGCGCAATTTATCATGA 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1648 GAGGTTTGGACGCTGGCCATTACAAAGCTGACGACCTGCGAAACTATCAAGGC 1589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 TATACATGCTCTCCGCTTACAGGTGACCAAAAGCAAGATGATTAATCCATTCAT 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1588 TACCACATGCTTTCGATTTAGCCGAGATCTTAAATGCGATGACATATGATTTAC 1529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 TTATTTATCAAAAGTCGCTGATACATGATGACAGCTGAAAAATGCTGGAAGATA 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1528 ATGTTCTATCAAAAGTCGCGCAAACTTCTATTTAGACAGTGAAGAAACGCTGCGCTC 1469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 TTTGAGATATGATTAATTTGTTCCAAATGATCCGATCTTAAATATCAAAACAGAG 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1468 TTTAAAGACAGCGCAAAATTCGATGCAAAATGATTTATCTTAAAGCCAAACAGAA 1409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 TGGTCAGTTCCTCTACTTTAAACCAAGATGCCAAGTCCGTTATTTCTATACAGATTAC 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1408 TGGTCAGTTCACACCATTTACATCTGACGAAAGAAATCCGTTATTTCTACATGATTTC 1349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 TCAGGTAATCTGTAAGATGGTGAAACCGGCTGCTGTAACCAAAATCAATTTCAACTGCTCAA 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1348 TCCGCTAAACATTA-----CGGCAAAACAAACACTGACACTGCACAA 1307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 GTAAACTTTCCAGCGGATGTCAGCTACACTTAAAGTCATGATGATTCATTA 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1306 GTTAAAGTATACAGATCAGACAGCTCT--TTGACATCAACGCTGAGGATTTATTA 1250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 TCTGTCTTGTATGCGGAGAGCTTACAGTTTATCAAAATTTATGCAATTTATGATGAA 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1249 TCAATCTTTGA---CGGTGACGGAAGAAACGTATCAAAATTTACAGCAGTTTCATCATGAA 1193
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 661 GCGAAGTGGATTTACAGTATAACATCTTTAAGAGACCCGATGATGTTGAAGATAG 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 1192 GGCACATACAGCTCAGCGCAACACCATACGCTGAGAGATCTTCCTACAGTAAGATPAA 1133
QY 721 GGCATTAATATCTTGTCTTTGAAGGAATACGTAAGCAACAGATGTTATCAAGCCGAT 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1132 GGCACAAATACTTAGATTGTAAGCAACACTGGAATGAAAGTGGTACCAAGCGCAA 1073
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 CACTCTTCAATATAAGCTTACTATAGCGGAGAGAGCTCTTCTTCCGAATGAAGAAA 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1072 GAATCTTTATTTAAACAAGCATCTATGCAAAAAGCAGATCTTCTCCGTCAAGAAAG 1013
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 AATTAACCTGCTTAAGCTCTTAAAAAACAAATTTCTTTTACGAATGGTGCATTAAGC 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1012 CAAAACTTCTGCAAAAGCGATTAATAAAAGCAGCGGTAGTTACCAAAAGCGGCTCTCGGT 953
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 ATGTGTAATTTGGCCCATGACTATACAGTAAAAAGTTTATGAAACCATTAAGTCGATCA 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 952 ATGATTTAGCTAAACGATGATTAACACTGTAATAAAAGTATGAACCGCTGATTCATCT 893
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 AACACAGTACAGATGATGATGCAAGCGGCCCATATATATTTAAATGAATTAATAGGTAT 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 892 AACACAGTACAGATGATGATGCAAGCGGCCCATATATATTTAAATGAACGGCAATGATAC 833
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1021 CTATTCAGGATTTCAAGAGATCCAAATGACAGATGATGAATTAACGCAAGATGTT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 832 CTGTTCACTGACTCCGCGGATCAAAAATGACGATTTGACGCGCATTTACGCTACAGATATT 773
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1081 TATATGCTAGGCGCCGAGCGGACTCTTAATGAGCCCAACACACCGGATTAATGAAGT 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 772 TACATGCTTGTATGATTTCTTAAATCTTTAACTGCGCCATTAACAGCGGCAACAAACT 713
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1141 GCACTTGTATTAACATGATCTTGAACCTGCTGATCTACACACACTTACTCTCATGTC 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 712 GCGCTTGTGTAAATAATGATCTTGAATCTTAAGATGATTAACCTTACTTACTACACTTC 653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1201 GGTATCCGCAACCTGAGGTAATATATGATGATCTCAAGATTAATGATGAGAGC 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 652 GCTGTACTCTCAAGCGAAGGAAACAAATGCTGATTTACAGCTATATGACAAACAGAGA 593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1261 TTTCTACCAAGATCATCTCTCAACCTGCGGAGACAAGCTTGGGTTATATTAAGGTCT 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 592 TTCTACGACAGCAACAAATCAATTAACGTTTGGCCCAAGCTTCTGCTGAACATCAAGGCAAG 533
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1321 GACACATCTGAGAGAGAAATAG 1343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 532 AAAACATCTGTTCAAGACAG 510
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: June 18, 2003, 07:25:31  
 Job time : 3564.48 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 02:50:32 ; Search time 1977.86 Seconds  
(without alignments)  
11177.167 Million cell updates/sec

Title: US-09-986-682B-4  
Perfect score: 1365  
Sequence: 1 ATGACACAGCGGAGCTACAA.....CCGACACAGACAAATTCACA 1365

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_estlum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359.6	26.3	755	17	B49063
2	309.6	22.7	623	17	A0284271
3	307.4	22.5	654	17	AG158197
4	296.8	21.7	687	17	AG145469
5	292.6	21.4	689	17	AG174787
6	290.8	21.3	653	17	A0311782

7	286.6	21.0	688	17	AG165661	AG165661 Pan trogl
8	284.4	20.8	656	17	AG158505	AG158505 Pan trogl
9	283.4	20.8	678	17	AG019951	AG019951 Homo sapi
10	271	19.9	711	17	AG001039	AG001039 Homo sapi
11	267	19.6	669	17	AG156831	AG156831 Homo sapi
12	264.8	19.4	581	17	AG0791204	AG0791204 HS_4555_B
13	258.4	18.9	682	17	AG020006	AG020006 Homo sapi
14	250	18.3	934	9	AL575209	AL575209 AL575209
15	248.6	18.2	686	17	AG185716	AG185716 Pan trogl
16	233	17.1	693	17	AG167579	AG167579 Pan trogl
17	228.6	16.7	622	17	AG160823	AG160823 Pan trogl
18	226	16.6	508	17	AG0317303	AG0317303 RPT11-10
19	220.8	16.2	1101	17	CNS00E5C	AL068761 Drosophila
20	215.8	15.8	488	17	AG0315317	AG0315317 RPT11-10
21	213	15.6	687	17	AG145995	AG145995 Pan trogl
22	194.6	14.3	661	17	AG019210	AG019210 Homo sapi
23	194.6	14.3	678	17	AG020005	AG020005 Homo sapi
24	192	14.1	698	17	AG165439	AG165439 Pan trogl
25	187.6	13.7	694	17	AG145903	AG145903 Pan trogl
26	186.6	13.7	385	17	B47667	B47667 RPT11-1A4
27	184.8	13.5	694	17	AG169167	AG169167 Pan trogl
28	183	13.4	682	17	AG165125	AG165125 Pan trogl
29	182.6	13.4	714	17	AG014172	AG014172 Homo sapi
30	179	13.1	674	17	AG165337	AG165337 Pan trogl
31	179	13.1	696	17	AG177482	AG177482 Pan trogl
32	173.6	12.7	729	17	AG014170	AG014170 Homo sapi
33	172.2	12.6	698	17	AG180444	AG180444 Pan trogl
34	170.8	12.5	601	17	P859L	AL390538 Leishmani
35	167.6	12.3	676	17	AG145781	AG145781 Pan trogl
36	167.2	12.2	705	17	AG147097	AG147097 Pan trogl
37	166.8	12.2	722	17	AG014192	AG014192 Homo sapi
38	165	12.1	683	17	AG168295	AG168295 Pan trogl
39	163.2	12.0	935	17	CNS008JC	AL052150 Drosophila
40	162.8	11.9	714	17	AG014191	AG014191 Homo sapi
41	162.4	11.9	672	17	AG149729	AG149729 Pan trogl
42	158.4	11.6	613	17	AG064045	AG064045 927P1-13C
43	154.6	11.3	752	17	AG171011	AG171011 Pan trogl
44	153.6	11.3	704	17	AG168501	AG168501 Pan trogl
45	151	11.1	731	17	AG162359	AG162359 Pan trogl

## ALIGNMENTS

RESULT 1  
B49063  
LOCUS RPT11-4121.TV RPT1-11 Homo sapiens genomic clone RPT1-11-4121, DNA  
DEFINITION  
B49063  
ACCESSION  
VERSION B49063.1 GI:2601300  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
1 (bases 1 to 755)  
Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden  
,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter  
,J.C.  
Use of BAC End Sequences for Sequence-Ready Map Building  
Unpublished (1997)  
CONTACT: Mark Adams  
DEPARTMENT OF Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdamas@tigr.org

Clones are derived from the human BAC library RPT1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from





[illegible][illegible]

FEATURES  
source

R.Site 1 : Ecori  
R.Site 2 : Ecori.  
Location/Qualifiers  
1. 687  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="RP43-006M11.T7"  
/sex="male"  
/cell\_type="lymphocytes"  
/clone.lib="RPCI-43 Chimpanzee Male BAC Library"

BASE COUNT 239 a 148 c 147 g 153 t

ORIGIN

Query Match 21.7%; Score 296.8; DB 17; Length 687;  
Best Local Similarity 68.7%; Pred. No. 4.7e-68;  
Matches 463; Conservative 0; Mismatches 187; Indels 24; Gaps 3;

QY 418 GAATGTCAGTCTCTGCTACTTTAACCAAGATGCCAGTCCGTTATTCTATACAGAT 477  
DB 30 GAATGTCAGTCTCTGCTACTTTAACCAAGATGCCAGTCCGTTATTCTATACAGAT 89  
QY 478 TACTCAGTAACTCTGAGATGGTGAACCGGTGCTGTAAACCAATCATTTCACTGCT 537  
DB 90 TATTCGGTAAACATTC-----GCAACCAAGCCTGACACAGCG 131  
QY 538 CAAGTAACTTATCCAGCGGATGAGTACCTTAAAGTCAGTATCTGATCAT 597  
DB 132 CAGTAAATGTGCAAAATCTGATGAC---ACACTCAAAATCAACGAGTGAAGATCAC 188  
QY 598 AATCTCTCTTATGCGGACGAGTACGATTTATCAAAATTTAGCAATTTATCAT 657  
DB 189 AAACAGATTTTGA---CGGAGACGGAACCAATCAGACGTTGACGATTTATCAT 245  
QY 658 GAAGCAAGTGAATTCAGTGTATTAACCATCTTTAAGACGCTCATCTATTTGAAGAT 717  
DB 246 GAAGCAATTAATCATTCGCGGACCAACCATACCTGAGAGACCTCATCTATTTGAAGAT 305  
QY 718 AAGGCCATTAATATCTTCTTCTTGAAGCGAATCTGGAACAACAGATGTTATCAAGC 777  
DB 306 AAGGCCATTAATATCTTCTTCTTGAAGCGAATCTGGAACAACAGATGTTATCAAGC 365  
QY 778 GATCAGTCTTCAATATAAGCTTACTATGCGGAGATGCTCTTCTTCCAGATGAA 837  
DB 366 GAAGCAATTAATATAAGCTTACTATGCGGAGATGCTCTTCTTCCAGATGAA 425  
QY 838 AAAAATAACTGCTTAAAGTCTTAAATAAACAATTTCTTTAGGAAATGTCATTA 897  
DB 426 AGCCAGAGCTTACGACGACGCTTAAATAAACAATTTCTTTAGGAAATGTCATTA 485  
QY 898 GGCATGTTGAATTTGGCCGATGACTATACAGTGAAGAGTGTATGAACCATTAATGCGCA 957  
DB 486 GGCATGTTGAATTTGGCCGATGACTATACAGTGAAGAGTGTATGAACCATTAATGCGCA 545  
QY 958 TCAACACAGTACGATGAGTGAAGCGGCAATATATTAAATGAATTAATATG 1017  
DB 546 TCAACACAGTACGATGAGTGAAGCGGCAATATATTAAATGAATTAATATG 605  
QY 1018 TATCTATTCAGGATTCAGAGATCCAAATAGCAGTGAATTAACGACAAAGAT 1077  
DB 606 TACTTGTTCAGTGAATTCAGCGGCTTCAAAATAGCAGTGAATTAACGACAAAGAT 665  
QY 1078 GTTATATGCTAGG 1091  
DB 666 ATTACATGCTTGG 679

RESULT 5  
AG174787 689 bp DNA linear GSS 09-JAN-2002  
LOCUS Pan troglodytes DNA, clone: RP43-045E23.T7, genomic survey  
DEFINITION  
sequence.  
ACCESSION AG174787  
VERSION AG174787.1 GI:16704467

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
PRIMERS

GSS.  
Pan troglodytes male lymphocytes DNA, clone.lib:RPCI-43 Chimpanzee  
Male BAC Library clone:RP43-045E23.T7.  
Pan troglodytes  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.  
1  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
BAC end sequences of library RPCI-43  
Direct Submission  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suenho-chou, Tsukuba, Ibaraki, Japan, 305-8565, Japan  
(E-mail:chimpes@sc.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the chimpanzee BAC library RPCI-43 This BAC  
end was generated during the R&D process and may have higher chance  
of clone tracking errors.

Sequencing: T7  
LIBRARY  
Vector : pBAC3.6  
R.Site 1 : Ecori.  
R.Site 2 : Ecori.  
Location/Qualifiers  
1. 689  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="RP43-045E23.T7"  
/sex="male"  
/cell\_type="lymphocytes"  
/clone.lib="RPCI-43 Chimpanzee Male BAC Library"

BASE COUNT 240 a 151 c 143 g 154 t 1 others

ORIGIN

Query Match 21.4%; Score 292.6; DB 17; Length 689;  
Best Local Similarity 68.9%; Pred. No. 6.1e-67;  
Matches 432; Conservative 0; Mismatches 189; Indels 6; Gaps 2;

QY 491 CTGAAGTGTGAGACCGGTGCTGTAACCAATCTTCACTGCTCAAGTAACTTAT 550  
DB 69 CTGACATTCCTGTAACATTTAGGCAACCAAGCTGACACGCGCATTAATGT 128  
QY 551 CCCAGCGGATGACGCTTAAAGTGAAGTATCTGATCTAATATCTGCTTGG 610  
DB 129 CAATATCTGAT--GACACATCAAAATCAAGGAGTGAAGATCAAAAGATTTTGG 185  
QY 611 ATGGCGGAGAGGTATAGTTTCAAAATATTCAGCAATTTATTCAGTGAAGCAAGTGA 670  
DB 186 A--CGGACACGGAACCAATATCAAGAGCTTCAAGATTTATTCAGTGAAGCAATTA 242  
QY 671 TTTCAAGTATACCATTTTAAAGAGACCTTCACTATTTGAATTAAGGCAATTAAT 730  
DB 243 CATCGCGACAAACCATTTAGGAGAGACCTTCACTATTTGAAGCAAGGCAATTAAT 302  
QY 731 ATCTTGTCTTTGAAGCAATTTAGGAGACCAAGATGTTATCAAGGCAATCTTGA 790  
DB 303 ACCTTGTATTCGAAGCAACGAGGAGCAAGAAAGGATGATCAAGATCTTTAT 362  
QY 791 ATAAATAGCTTACTTGGCGAGGAGTGTCTTCTCCAGATGAAGAAATTAATCTGC 850  
DB 363 TTAAACAGCGTACTAGCGGCGGCGGACGACATCTTCTCCAGATGAAGCAAGATCTGC 422  
QY 851 TTCAAAGTCTAAACCAATTTCTTTAGGCAATTTGATGATGATGATGATGAT 910  
DB 423 AGCAGAGCGCTAAACCAAGCATGTGATTTAGCAGAGCGGCGGCTGATCATAGAT 482  
QY 911 TGCGCGATGACTATACGTAAGAAAGTGTATGAACCAATTAAGTGCATCAACACAGTAG 970

Db	483	TAATATGATTAACCACTTGAAGAAAAAGTAATGAAGCCCCCTGATCTTCAACAGGTGA	542
OY	971	CAGATGAAGTGGAAAGCGCCCAATATATTAATTAATGAATATTAATGTATCTAATCACGG	1036
Db	543	CTGATGAAATGACAGCGCCCGAATGTTTTCAAATGAATGACCAAAATGATCTTGCTACATG	602
OY	1091	ATTCAGAGATTCGCAAAATGACGAGTGATGTAATTAACGACAAAGATCTTATATAGCTAG	1099
Db	603	ATTCAGCGGGTTCGCAAAATGACGATGATGATTAATTAACGATATTAATACAGTTCGCTTG	662
OY	1091	GGCCCGAGGCGACTCCTTTAATGACC	1117
Db	663	GTTATGTTATCAAACTCTTAACCGCC	689
RESULT 6			
LOCUS	AQ311782	653 bp	DNA
DEFINITION	RPC111-103A24.TV	RPC1-11	Homo sapiens genomic clone RPC1-11-103A24
ACCESSION	AQ311782		
VERSION	AQ311782.1	GI:4043531	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 653)		
TITLE	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.		
JOURNAL	Use of human BAC End sequences for Sequence-Ready Map Building Unpublished (1998)		
COMMENT	Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbe@tigr.org Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu">http://bacpac.med.buffalo.edu</a> ) or from Research Genetics ( <a href="http://inforesgen.com">inforesgen.com</a> ). BAC end search page: <a href="http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html">http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html</a> Seq primer: 77 Class: BAC ends.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..653		
	/organism="Homo sapiens"		
	/db_xref="GDB:7539191"		
	/db_xref="taxon:9606"		
	/clone="RPC1-11-103A24"		
	/clone_lib="RPC1-11"		
	/sex="Male"		
	/cell_type="Lymphocytes"		
	/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC library"		
BASE COUNT	232 a	145 c	143 g
ORIGIN	133 t		
Query Match	21.3%	Score 290.8;	DB 17; Length 653;
Best Local Similarity	68.0%;	Pred. No. 1.8e-66;	
Matches 460;	Conservative 0;	Mismatches 192;	Indels 24; Gaps 3
OY	331	ATTACAGAGCTGGAAAAATGCTGGGAAGATATTGACAGTATGATGAATAATTGGTTCCAAT	390
Db	2	ATCAGACAGCTGGAAAAAGCGGCGCTGCTTTAAAGACAGCGATTAAGTTGACGCCAAC	61
OY	391	GATCCGATCTTAAATATCAACACAGAGAGGTCAGGTTGCTGACTTTTAACCAAAAT	450
Db	62	GATCCGATCTCGAAGATCACACGCAAGATGTCGGTTTCGCAACTTTTACATCTTAC	121

OY		451	GGCCACAGTCCTTATTCTGTATACAGCATTTACTCAGGTAATCCTGAAGAATGGTGAAACCGGT	510
Db		122	GGAAAAATCGTTTATTTCTTACACGTACATTTTCGGTAAACAFTA-----	165
OY		511	GCTGGTAAACCATAATCTATTTCAACAGTCCTCAAGTAAACCTTATCCAGCCGGATGCATCA	570
Db		166	--CGGCAACAACAGGCTCACCAACAGCGGAGTAAATGTGTCAAATCTGAT---GACACA	220
OY		571	CTTAAAGTCGATGGAGTATCTGATCATAAATCTGTCTTTGATGGCGGAGACGGTACAGTT	630
Db		221	CTCAAAAATCAACGGAGTGGAGATTCACAAAACGATTTTTTGA---CGAGACGGAACAAACA	277
OY		631	TATCAAAATATTCCAGCATTTTATCGATGAAGGCAAGTGGATTTCCAGTGATATACCATCT	690
Db		278	TATCAGAAGCTTCAGCAGTATTTATCGATGAAGGCAATTTATCATCTCGGGACACACCATAGC	337
OY		691	TTAAGAGACCCCTCACTATGTGAAGATPAAGGGCCATAATATCTGTCTTTGAAGCGAAT	750
Db		338	CTGAGAGACCCCTCACTACGTGTGAAGSACAAAGGCCATTAATACCTGTATTTGGAAGCCAC	397
OY		751	ACTGGAACACAGATGTTATTCAGAGCGATCAGTCTTTCAATPATTAAGCTTACTATGGC	810
Db		398	ACGGGAACAGAAAACGATACCAAGGCGAAMAATCTTTATTTAACAAACCGCTACTACGCG	457
OY		811	GGAGTAGACGTCCTCTTCCAGATCAAAAAAATAACGTCCTCAAAAGTCCTAAAAAACAA	870
Db		458	GGCGGCACGAACTCTCTTCGTTAAAGAACCCAGAAAGCTTTCAGCAAGCGCTRAAAAAAGC	517
OY		871	ATTGCTCTTATAGCGAATGTGCATATTAGGCATTTGTTGAATTGGCCGATGACTATACAGTG	930
Db		518	GATGCTGATGTTAGGSAACGGGCGCCCTCGTGTATCAGATTAATAATATGATTACCAATTG	577
OY		931	AAAAGTGTATTGAACACATTAGTGCATCAACACAGTACGATGAAGTGAAGCGCGC	990
Db		578	AAAAAGATTAAGAAAGCCGCTGATCATCTTTATACAGGTAAGTGAAGTAATGAGCGCGC	637
OY		991	AATATATTTAATAATGA	1006
Db		638	AATGTTTCCAANAATGA	653
RESULT 7				
LOCUS	AG165661	688 bp	DNA	linear
DEFINITION	Pan troglodytes DNA, clone: RP43-033D03.T7, genomic survey sequence.			GSS 09-JAN-2002
ACCESSION	AG165661			
VERSION	AG165661.1			GI:16695339
KEYWORDS	GSS:			
SOURCE	Pan troglodytes male lymphocytes DNA, clone_1lb:RCI-43 Chimpanzee Male BAC library clone:RP43-033D03.T7.			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.			
REFERENCE	1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.			
AUTHORS	BAC end sequences of Library RCI-43 Unpublished			
JOURNAL	2 (bases 1 to 688)			
REFERENCE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.			
AUTHORS	Direct Submission			
TITLE	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-Chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbeg@sc.riken.go.jp, URL:http://npg.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)			
JOURNAL	Clones are derived from the chimpanzee BAC library RCI-43 This BAC end was generated during the Rd process and may have higher chance of clone tracking errors.			
COMMENT	PRIMERS			



Db 479 TAAATATGATTAACATTTGAAAAAGTAATGAAGCGGTGATCATTCAACACGGTAA 538

OY 971 CAGATGATGTCGACGCGCCATATATTAATGAATATTAATGATGATCTATTCGCG 1030

Db 539 CTATGATTAATCGACCGCGCAATGTTTCAAAATGAACGCAATGATCTGTCCTG 598

OY 1031 ATTCAAGAGATCCAAAATGACGATGATGAATTAACGACAAAAGATGTTATATGCT 1088

Db 599 ATTACGCGGTTCAAAAATGACATGATGATTTAACTCAAAAGATTTTACATGCT 656

RESULT 9  
AG019951/c  
LOCUS  
DEFINITION  
AG019951 678 bp DNA linear GSS 16-OCT-1999  
Homo sapiens genomic DNA, 21q region, clone: B293N7 H058(RP),  
genomic survey sequence.  
ACCESSION  
AG019951 GI:6045895  
VERSION  
AG019951.1 GI:6045895  
KEYWORDS  
Homo sapiens DNA, clone:B293N7 H058(RP).  
SOURCE  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 678)  
Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,  
Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.  
Homo sapiens genomic DNA, chromosome 21q  
Published only in Database (1999)  
2 (bases 1 to 678)  
Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,  
Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.  
Direct Submission  
Submitted (16-OCT-1999) Masahira Hattori, RIKEN Genomic Sciences  
Center(GSC) C/O Kitasato University, 1-15-1 Kitasato, Sagamihara  
228-8555, Japan (E-mail:hattori@hg.ims.u-tokyo.ac.jp,  
Tel:042-778-9923, Fax:042-778-9924)  
LOCATION/Qualifiers  
FEATURES  
source  
1..678  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="21q  
/clone="B293N7 H058(RP)"  
BASE COUNT 140 a 153 c 154 g 231 t  
ORIGIN

Query Match 20.8%; Score 283.4; DB 17; Length 678;  
Best Local Similarity 68.2%; Pred. No. 1.7e-64;  
Matches 463; Conservative 0; Mismatches 191; Indels 25; Gaps 4;

OY 171 AATGATTTAGATGTATGGATAGCTGGCCACGCAAAAGCTGATCTGCGGCAAA 230

Db 656 AAAAGAGCTTGATGTGGGACAGCTGGCGGTGCAAAAGCTGACGACAGTATGTA 597

OY 231 TTATCATGATATACATGCTCTCCGCTTATGACGATGACCAAAAACAGTATGTA 290

Db 596 ATCAACGGGTATCAGTGTGTGTTGCTCTGGGGAAGCGCAAGACCTGATGACAC 537

OY 291 TCCACTTCATTTATTTATCAAAAAGTGGTGTATCATGCTGACGCTGG-AAAAAG 349

Db 536 ATCAATCTCATGCTTTTATCAAAAGTGGCGACAACTCAATGACAGTGGAAAAAGC 477

OY 350 CTGGAAGATATTTGAAGATATGATTAATTTGTTCCAAATGATCCGTATCTTAATAT 409

Db 476 CCGGCGGTGCTTTAAAGACAGGATTAAGTTGACGCGCAATCCGATCTAAAGATC 417

OY 410 AAACACAGAGTGTGAGGTTCTGCTATCTTAACCAAGATGCGCAAGTCTTTATTC 469

Db 416 AGACGCAAGATGTCGCGGTGCAACCTTACATCTGACGAAAAATCCGTTATCT 357

OY 470 ATACATTTACTAGTATCTCTGAAGATGTTGGAACCGGTGCTGTTACCAATATCT 529

Db 356 ACACGTACTATTCGGGTAAACATTA-----CGGCAACAAAGCTGA 315

OY 530 CAACGTCTCAAGTAACTTATCCACGCGGATGACAGTACACTTAAGTATGAGTAT 589

Db 314 CAACAGCGCAGGTAAATGCTCAAAATCTGAT---GACACATCAAAAATCAACGAGTG 258

OY 590 CTGATCATTAATCTGCTTTGATGCGGAGACGGTACGTTATCAAAATATTCACCAAT 649

Db 257 AAGATCACAAAAGATTTTTGA---CGGACAGGAAAAACATATTCGAAGCTTCACAGT 201

OY 650 TTATCATGTAAGCGCAAGTGAATTTACGATGATACCATCTTTAAGACCTCTACTATG 709

Db 200 TTATCGATGAAGCAATTTATATACATCCGCGCAACCATACGCTGAGAGACCTCTAC 141

OY 710 TTGAAGATGAAGCGCAATATCTGCTTTGAAGCGAATTAAGCAATGATGATGATG 769

Db 140 TTGAAGCAAAAGCGCATTAATACCTGTATTCGAAGCAACAGGGAAGAAACGAT 81

OY 770 ATCAAGCGATCAAGTCTTTCAATATTAAGCTACTATGCGGGAAGTACGCTCTTC 829

Db 80 ACCAAGCGCAAGATCTTTATTAACAACGCTACTAGCGGCGGCGCAGCAACTTCTCC 21

OY 830 AGAATCAAAAATTAATCT 848

Db 20 GTAAGAAAGCCAGAACT 2

RESULT 10  
AG001039/c  
LOCUS  
DEFINITION  
AG001039 711 bp DNA linear GSS 06-FEB-1999  
Homo sapiens genomic DNA, 21q region, clone: S39BG29, genomic  
survey sequence.  
ACCESSION  
AG001039 GI:2579847  
VERSION  
AG001039.1 GI:2579847  
KEYWORDS  
GSS.  
Homo sapiens DNA, clone:S39BG29.  
SOURCE  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 711)  
Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.  
Homo sapiens genomic DNA, chromosome 21q  
Published only in Database (1997)  
2 (bases 1 to 711)  
Hattori,M., Ishii,K., Shiba,T. and Sakaki,Y.  
Direct Submission  
Submitted (31-OCT-1997) Masahira Hattori, Kitasato University,  
Department of Science, JST Sequencing Laboratory, Kitasato 1-15-1,  
Sagamihara 228, Japan (E-mail:hattori@hg.ims.u-tokyo.ac.jp,  
Tel:0427-78-9732, Fax:0427-78-9561)  
LOCATION/Qualifiers  
FEATURES  
source  
1..711  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q"  
/clone="S39BG29"  
BASE COUNT 145 a 158 c 166 g 241 t 1 others  
ORIGIN

Query Match 19.9%; Score 271; DB 17; Length 711;  
Best Local Similarity 67.3%; Pred. No. 3.5e-61;  
Matches 466; Conservative 0; Mismatches 201; Indels 25; Gaps 5;

OY 328 TCGATTGACAGCTGGAATAAGTGTGAAGATTTGAAGATATGTAATTTGTTCCA 387

Db 708 TCAATGACAGCTGGAATAAGCGCGCGCTGCTTTAAGACAGATAGTTGACAGCCC 649

OY 388 AATGATCCGATATCTTAATTAACAACAGAGATGCTGAGTCTGCTACTTAACCCAA 447

Db 648 AAGGATCCG-ATCTGAAGATCAAGCA-GAATGGTCCGGTCTGCAACCTTTACATCT 591

OY 448 GATGCCAAGTCCGTTATCTTATACAGATTAATCTAGATTAATCTGAAGATGGTGAAC 507

Db 590 GACGGAAAAATCCGTTTATTTCTACACTGACTATTCGCGTAACAT----- 546  
 QY 508 GGTGCTGTAAACCAATCATTTTCACTGCTCAAGTAATTAATCCAGCCGAGTGCAGT 567  
 Db 545 --TACGCCCAACAAAGCCCTGACACAGCCAGGTAATATGTCAAAATCTGAT---GAC 491  
 QY 568 ACACTTAAAGTCAGTAGAGTATCTGATCAATATCTGCTTTGATGGCGAGACGGTACA 627  
 Db 490 ACACTCAAAATCAACGAGTGAAGATCACAAAACGATTTTGA---CGGAGACGGAAAA 434  
 QY 628 GTTATCAAAATATTCGCAATTTATCGAATGAAGCGCAAGGATTTAGGTGATACAT 687  
 Db 433 ACATATCAAGACGTTTACGCAATTTATCGATGAAGCAATTAATACATCCGCGACAAACAT 374  
 QY 688 ACTTTAAGACACCTCCTCATGTTTGAAGATAGGCGCATTAATATCTGCTTTGAAGCG 747  
 Db 373 ACGGTGAGACACCCNTACTACGTTTGAAGACAAAGCCATTAATATCTGATTTGCAAGCC 314  
 QY 748 AATACCTGAACAAACAGATGTTATCAAGCGGATCAGTCTTTCAATTAATAAAGCTTACTAT 807  
 Db 313 AACACGGGAACAGAAAACGGATACACAGCGCAAGAAATCTTATTTAACAAGCGTACTAC 254  
 QY 808 GGGGAGATGACGCTCTTCCAGAAATGAATAAATTAAGTCTTCAAGTCTTAATAA 867  
 Db 253 GGGCGCGGACGAACTTCTTCCGTAAGAAAGCCAGAGCTTCAAGCGCGCTTAATAA 194  
 QY 868 CAAATTCCTCTTGAAGCAATGTCATAGGCAATTTGATTTGCGGATGACTATACA 927  
 Db 193 CGGATGCTGATTTAGGAACGGCGCCCTCGGTATCTATAGATTAAATATGATTTACACA 134  
 QY 928 GTGAAAGCTTTATGAACCATTTAGTCGATCAACACAGTAGAGATGAAGTGAAGCG 987  
 Db 133 TTGAAAAAGTATGAAGCGCGTGCATCTTCAAAACGCGTAACCTGTAATGAAATCGAGCG 74  
 QY 988 GCCAATATTTTAAATGAATTAATTAATGTA 1019  
 Db 73 GCCAATTTTTCAAAATGAACGCCAATGTA 42  
 RESULT 11  
 AG156831  
 LOCUS  
 DEFINITION  
 AG156831 669 bp DNA linear GSS 09-JAN-2002  
 sequence.  
 AG156831  
 ACCESSION  
 AG156831.1 GI:16686509  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pan troglodytes male lymphocytes DNA, clone: RP43-021J04.T7.  
 Male BAC Library clone:RP43-021J04.T7.  
 GSS.  
 Pan troglodytes male lymphocytes DNA, clone:lib:RP43-43 Chimpanzee  
 Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 REFERENCE  
 AUTHORS  
 1  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE  
 BAC end sequences of library RP43-43  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 669)  
 TITLE  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 JOURNAL  
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:chimbseq@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/  
 tel:81-45-503-9111, Fax:81-45-503-9170)  
 clones are derived from the chimpanzee BAC library RP43-43 This BAC  
 end was generated during the R&D process and may have higher chance  
 of clone tracking errors.  
 PRIMERS  
 Sequencing: T7  
 LIBRARY  
 Vector : pBACe3.6  
 COMMENT

FEATURES  
 source  
 R Site 1 : EcorI  
 R Site 2 : EcorI.  
 location/Qualifiers  
 1. 669  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9598"  
 /clone="RP43-021J04.T7"  
 /sex="male"  
 /cell\_type="lymphocytes"  
 /clone\_lib="RP43-43 Chimpanzee Male BAC Library"  
 BASE COUNT  
 227 a 153 c 143 g 146 t  
 ORIGIN  
 Query Match 19.6%; Score 267; DB 17; Length 669;  
 Best Local Similarity 65.5%; Pred. No. 4e-60;  
 Matches 390; Conservative 0; Mismatches 205; Indels 0; Gaps 0;  
 QY 749 ATACTGGAACACAGATGTTATCAAGCGCATGCTTTCAATTAATAAAGCTTACTATG 808  
 Db 34 ACACGGGAACGAAAACGGATACCAAGCGCAAGAACTTTATTTAACAAGCGTACTACG 93  
 QY 809 GCGAAGTGACGTCCTTCTCCAGATGAATAAATAACTGCTTCAAGTCTTAATAAAC 868  
 Db 94 GCGCGGACGAACTTCTTCCGTAAGAAAGCCAGAAAGCTTCAAGCGCGCTTAATAAAC 153  
 QY 869 AATTCCTCTTGAAGCAATGTCATAGGCAATTTGTAATTTGCGGATGACTATACAG 928  
 Db 154 GCGATGCTGATTTAGGAACGGCGCCCTCGGTATCTATAGTTAAATATGATTTACAT 213  
 QY 929 TGAAGAAGTGTATGAACCATTTAGTCGATCAACACAGTAGAGATGAAGTGAAGCGG 988  
 Db 214 TGAATAAGTAAATGAAGCCGCTGATCATCTTCAAAACGCGTAACGTGAATCGACGCG 273  
 QY 989 CCAATATTTTAAATGAATTAATTAATGATATTTCAAGATTCAGAGATGCCAAAA 1048  
 Db 274 CGATGTTTCCAAATGAAGCGCAATGACTTTGTCAGTATTCAGCGGTTCAAAA 333  
 QY 1049 TGACAGATGATGAATTAACGACAAAGATTTTATGCTAGAGCCCGGAGCGACTCT 1108  
 Db 334 TGACGATGATGATGATTAATTAACGATATTTACATGCTGTTATGATCAAACTCT 393  
 QY 1109 TAAATGCCGCCACCAACCCGATTAATGAACATGAGTCTTATGAACATGATCTTGACC 1168  
 Db 394 TAACCGGCCCTTCAAGCCGCTGTAACAAAACAGCGCTTGTGCGAATAAAGGCTTGATC 453  
 QY 1169 CTGCTGATTCACACACACTTACTCTCATTTGCGGTATCCGCGACCTGAAGTAAATAG 1228  
 Db 454 CAACGATGATGATTTACTTACTCTCTCATCTTCCAGAGTCCGCAAGCAAGCAACAAATG 513  
 QY 1229 TGTACTACACAGTTATATGACGAATAGAGGCTTTCTATCCAGAACATCACTCTACCTGC 1288  
 Db 514 TGTATTACAAACCTACATGACAAACAGAGGCTTTCTGAGATAAAGGCAACATTTG 573  
 QY 1289 GGCACAAGCTTGGGTTAATTAATGAAGGCTGACATCTGAGAGAGAAATAG 1343  
 Db 574 GCGCAAGCTTTTAATGAACATCAAGGCAATTAACATCCGTTGCAAAAACAG 628  
 RESULT 12  
 AO791204  
 LOCUS  
 DEFINITION  
 AO791204 581 bp DNA linear GSS 03-AUG-1999  
 HS\_4555\_B1.A10.T7A CIT Approved Human Genomic Sperm Library D Homo  
 sapiens genomic clone Plate=4555 Col=19 Row=B, DNA sequence.  
 AO791204  
 ACCESSION  
 AO791204.1 GI:5698916  
 VERSION  
 KEYWORDS  
 GSS.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 581)  
 Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T.,  
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D. and



QY 1252 AATAGAGGCTTATCCAGAACATCATCTCAGCTGCGGAGAACAGCTGGGTTATATAT 1311  
 DB 201 AACAGAGGCTTCTTCGAGGATTAAGGACACATTTTCAAGCTTTAATGAACATC 142  
 QY 1312 AAAGGCTGACATCTGAGAGAGAAATAG 1343  
 DB 141 AAAGGCAATTAACATCCGTTCTCAAAACAG 110  
 RESULT 14  
 AL575209 934 bp mRNA linear EST 16-FEB-2001  
 LOCUS AL575209 LTI.NFL006.PL2 Homo sapiens cDNA clone CS0D1062YP02 5  
 DEFINITION prime, mRNA sequence.  
 ACCESSION AL575209  
 VERSION AL575209.1 GI:12936153  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 934)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
 FEATURES  
 source  
 1. 934  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1062YP02"  
 /clone.lib="LTI.NFL006.PL2"  
 /tissue\_type="placenta"  
 /note="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang life technologies, a division of invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@life.com URL : http://fulllength.invitrogen.com"  
 BASE COUNT 298 a 196 c 223 g 209 t 8 others  
 ORIGIN  
 Query Match 18.3%; Score 250; DB 9; Length 934;  
 Best Local Similarity 64.8%; Pred. No. 1.5e-55;  
 Matches 367; Conservative 2; Mismatches 197; Indels 0; Gaps 0;

QY 1078 GTTATATATGAGGCGCGAGCGATCTCTTAATATGCGCACACACCGATTAATGAA 1137  
 DB 323 ATTATATGCTTGTATGATATCAACACTTAAACGGCGCTTACAAAAMAAAAA 382  
 QY 1138 ACTGACCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1197  
 DB 383 AAAGGCTGCTGCTGCAAAATGGGCTTGTATGATGATGATGATGATGATGATGAT 442  
 QY 1198 TCGGATATCCGCGCGCGAGCGATCTCTTAATATGCGCACACACCGATTAATGAA 1257  
 DB 443 TTGCGAGTCCCGCGCGAGCGATCTCTTAATATGCGCACACACCGATTAATGAA 502  
 QY 1258 GCGCTTATATGAGGCGCGAGCGATCTCTTAATATGCGCACACACCGATTAATGAA 1317  
 DB 503 GCGCTTATATGAGGCGCGAGCGATCTCTTAATATGCGCACACACCGATTAATGAA 562  
 QY 1318 TCTGACATCTGAGAGAGAAATAG 1343  
 DB 563 AATAAACAATCCGTTCTCAAAACAG 588  
 RESULT 15  
 AG185716 696 bp DNA linear GSS 09-JAN-2002  
 LOCUS AG185716  
 DEFINITION Pan troglodytes DNA, clone: RP43-059006.T7, genomic survey  
 sequence.  
 ACCESSION AG185716  
 VERSION AG185716.1 GI:16715396  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes male lymphocytes DNA, clone.lib:RP43-Chimpanzee  
 Male BAC library clone:RP43-059006.T7.  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.  
 REFERENCE 1  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Toki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE BAC end sequences of library RP43-43  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 696)  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Toki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suenho-chou, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:chiimpes@sc.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 COMMENT Clones are derived from the chimpanzee BAC library RP43-43 This BAC  
 end was generated during the R&D process and may have higher chance  
 of clone tracking errors.  
 PRIMERS  
 Sequencing: T7  
 LIBRARY  
 Vector : pBACE3.6  
 R.site 1 : EcoRI  
 R.site 2 : EcoRI.  
 Location/Qualifiers  
 1. 696  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9598"  
 /clone="RP43-059006.T7"  
 /sex="male"  
 /cell\_type="lymphocytes"  
 /clone.lib="RP43-Chimpanzee Male BAC Library"  
 BASE COUNT 237 a 162 c 148 g 146 t 3 others  
 ORIGIN  
 Query Match 18.2%; Score 248.6; DB 17; Length 696;  
 Best Local Similarity 64.6%; Pred. No. 3.2e-55;  
 Matches 459; Conservative 0; Mismatches 200; Indels 51; Gaps 4;







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 07:25:42 ; Search time 206.939 Seconds  
(without alignments)  
9679.375 Million cell updates/sec

Title: US-09-986-682B-4  
Perfect score: 1365  
Sequence: 1 ATGACAGCGCGGACATACAA.....CCGACAGACAGCAATTCACA 1365

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEM\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCRUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1365	100.0	1365	10	US-09-986-682B-4
2	1365	100.0	2408	10	US-09-986-682B-5
3	617.2	45.2	1449	10	US-09-974-300-2644
4	578.6	42.4	2253	10	US-09-993-292A-19
5	578.6	42.4	8908	10	US-09-993-292A-18
6	577	42.3	4938	9	US-10-055-794-1
7	577	42.3	6525	9	US-10-055-794-3
8	577	42.3	10317	10	US-09-849-866-1
9	135.6	9.9	368	10	US-09-864-761-21746
10	73.2	5.4	376	10	US-09-864-761-5018
11	72.4	5.3	1146	10	US-09-778-844-136
12	61.2	4.5	4634	10	US-09-995-587A-10
13	52	3.8	2367	10	US-09-995-587A-2
14	52	3.8	2394	10	US-09-995-587A-3
15	52	3.8	2592	10	US-09-995-587A-4
16	44	3.2	58985	9	US-09-901-152-3
17	41.8	3.1	5301	7	US-08-781-986A-443
18	38.6	2.8	921	10	US-09-815-242-4588
19	38.6	2.8	957	10	US-09-815-242-8609

C 20	38.4	2.8	5086	9	US-09-764-891-9794	Sequence 9794, Ap
21	38	2.8	314	10	US-09-833-381-190	Sequence 190, Ap
22	38	2.8	548	10	US-09-874-300-3392	Sequence 3392, Ap
23	37.8	2.8	3831	12	US-10-051-952-5	Sequence 5, Ap
24	37.4	2.7	8761	9	US-10-098-841-11	Sequence 11, Ap
25	37.4	2.7	10758	12	US-10-044-090-61	Sequence 61, Ap
26	37.2	2.7	3996	6	US-10-087-464-42	Sequence 42, Ap
27	37	2.7	1799	7	US-08-781-986A-447	Sequence 447, Ap
28	36.6	2.7	7101	9	US-10-011-366-9	Sequence 9, Ap
29	36	2.6	1795	10	US-09-764-864-145	Sequence 145, Ap
30	36	2.6	2971	10	US-09-815-242-3842	Sequence 3842, Ap
31	36	2.6	5136	10	US-09-070-927A-270	Sequence 270, Ap
32	35.6	2.6	556	10	US-09-864-761-8619	Sequence 8619, Ap
33	35.4	2.6	439	10	US-09-864-761-20174	Sequence 20174, A
34	35.4	2.6	3317	9	US-10-055-412B-1	Sequence 1, Ap
35	35.4	2.6	378361	9	US-09-901-136-3	Sequence 3, Ap
36	35	2.6	327	10	US-09-864-761-28059	Sequence 28059, A
37	35	2.6	456	10	US-09-864-761-11468	Sequence 11468, A
38	35	2.6	2058	9	US-10-216-289-3	Sequence 3, Ap
39	35	2.6	14483	9	US-09-818-657-3	Sequence 3, Ap
40	35	2.6	82938	9	US-09-864-761-2617	Sequence 2617, Ap
41	34.8	2.5	470	10	US-10-123-155-192	Sequence 192, Ap
42	34.6	2.5	513	9	US-09-938-842A-4312	Sequence 4312, Ap
43	34.6	2.5	772	9	US-10-239-676-212	Sequence 212, Ap
44	34.6	2.5	8996	9	US-10-171-311-7	Sequence 7, Ap
45	34.6	2.5	12313	9		

#### ALIGNMENTS

RESULT 1  
US-09-986-682B-4  
Sequence 4, Application US/09986682B  
Patent No. US20020115182A1  
GENERAL INFORMATION:  
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU  
KENKYUJO  
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEWMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/986, 682B  
FILING DATE: 06-NO. US20020115182A1-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 170,630/1996  
FILING DATE: 10-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TSUSAKI-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-737-3528  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ. ID NO. 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1365 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-986-682B-4

Query Match 100.0%; Score 1365; DB 10; Length 1365;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGAACACGGGGGACTACAGAGACATGTTTGGCCATTTATACACGGCTGACATG 60
DB 1 ATGAACACGGGGGACTACAGAGACATGTTTGGCCATTTATACACGGCTGACATG 60
OY 61 CTAAAAATTCAGAGACAAACAACAGTCTCAATTTAAAGTCCCTCAATTCATGATCA 120
DB 61 CTAAAAATTCAGAGACAAACAACAGTCTCAATTTAAAGTCCCTCAATTCATGATCA 120
OY 121 GCATCAAAAACATTTGATGGGCAAAAGGATGATTAAGTACAGCACTTATGATTTA 180
DB 121 GCATCAAAAACATTTGATGGGCAAAAGGATGATTAAGTACAGCACTTATGATTTA 180
OY 181 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
DB 181 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
OY 241 TATACATCTGCTCGCTTTAGACAGGTGACCCAAAAACAGTATGATGATGATGATG 300
DB 241 TATACATCTGCTCGCTTTAGACAGGTGACCCAAAAACAGTATGATGATGATGATG 300
OY 301 TTATCTCTCAAAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
DB 301 TTATCTCTCAAAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
OY 361 TTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 361 TTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
OY 421 TGGTACAGTCTGCTGCTTTAACAAGATGATGATGATGATGATGATGATGATGATG 480
DB 421 TGGTACAGTCTGCTGCTTTAACAAGATGATGATGATGATGATGATGATGATGATG 480
OY 481 TCAGATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
DB 481 TCAGATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
OY 541 GTAACCTTATCCAGCCGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
DB 541 GTAACCTTATCCAGCCGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
OY 601 TCTGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
DB 601 TCTGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
OY 661 GGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
DB 661 GGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
OY 721 GGCATTAATATCTGCTTTGAAGCGAATCTGATGATGATGATGATGATGATGATGATG 780
DB 721 GGCATTAATATCTGCTTTGAAGCGAATCTGATGATGATGATGATGATGATGATGATG 780
OY 781 CAGCTTTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
DB 781 CAGCTTTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
OY 841 AATAAAGCTGTTCAAAAGCTTAAAAAACAAATGCTTTTAAAGGAATGGTCAATTAGG 900
DB 841 AATAAAGCTGTTCAAAAGCTTAAAAAACAAATGCTTTTAAAGGAATGGTCAATTAGG 900
OY 901 ATTGTTGAATGGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
DB 901 ATTGTTGAATGGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
OY 961 AACACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
DB 961 AACACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020

```

```

DB 961 AACACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
OY 1021 CTATTCACGGATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
DB 1021 CTATTCACGGATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
OY 1081 TATATGCTAGGGGCGGAGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
DB 1081 TATATGCTAGGGGCGGAGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
OY 1141 GCACTTGTATGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
DB 1141 GCACTTGTATGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
OY 1201 GGTATCCCGCACCCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
DB 1201 GGTATCCCGCACCCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
OY 1261 TTCTATCCAGAACATCACTCTCACCTCGGACACAGCTTGGGGTTAATATTAAGGGTCT 1320
DB 1261 TTCTATCCAGAACATCACTCTCACCTCGGACACAGCTTGGGGTTAATATTAAGGGTCT 1320
OY 1321 GACACATCTGGAGAGAAATAGTTCCGGACACAGACAAATTCCTCA 1365
DB 1321 GACACATCTGGAGAGAAATAGTTCCGGACACAGACAAATTCCTCA 1365

```

# RESULT 2

US-09-986-682B-5  
Sequence 5, Application US/09986682B

Patent No. US20020115182A1

GENERAL INFORMATION:

APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU

TITLE OF INVENTION: POLYPEPTIDE HAVING - FRUCTOPURANOSIDASE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/986, 682B

FILING DATE: 06-NO. US20020115182A1-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 170, 630/1996

FILING DATE: 10-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25, 618

REFERENCE/DOCKET NUMBER: TSUSAKI-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

SEQUENCE CHARACTERISTICS:

LENGTH: 2408 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORGANISM: Bacillus sp.

INDIVIDUAL ISOLATE: V230 (FERM BP-5054)

FEATURE:  
 NAME/KEY: signal peptide  
 LOCATION: 1..360  
 IDENTIFICATION METHOD: E  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
 US-09-986-682b-5

Query Match 100.0%; Score 1365; DB 10; Length 2408;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAACAGCGGGAGCTACAGAGAGACTATGTTGGCCATATTACAGCGCTGACATG 60
DB 457 ATGAACAGCGGGAGCTACAGAGAGACTATGTTGGCCATATTACAGCGCTGACATG 516
QY 61 CTAATAATTCAGAGCAACAACAAAGCTCCTAATTAAAGGCCCATTCATCAATGCATCA 120
DB 517 CTAATAATTCAGAGCAACAACAAAGCTCCTAATTAAAGGCCCATTCATCAATGCATCA 576
QY 121 GCAATCAAAAACATTTGATTCGGCAAAAAGGATATGTAAGTCAGCACTTAATAGATTYA 180
DB 577 GCAATCAAAAACATTTGATTCGGCAAAAAGGATATGTAAGTCAGCACTTAATAGATTYA 636
QY 181 GATGATGGGATAGCTGGCCACTGCAAAAACGCTGATGCTACTCGCGCAATTTATCATGA 240
DB 637 GATGATGGGATAGCTGGCCACTGCAAAAACGCTGATGCTACTCGCGCAATTTATCATGA 696
QY 241 TATCAGATCGTCTCGGCTTTAGCAGGTGACCAAAAACAGTATGATATCCCATTTAT 300
DB 697 TATCAGATCGTCTCGGCTTTAGCAGGTGACCAAAAACAGTATGATATCCCATTTAT 756
QY 301 TTATTCATCAAAAAGTCGGTGTATACATGATTTGACAGCTGGAAAAATGCTGGAAGACTA 360
DB 757 TTATTCATCAAAAAGTCGGTGTATACATGATTTGACAGCTGGAAAAATGCTGGAAGACTA 816
QY 361 TTTGACATATGATTAATTTGTTCCAAATGATCCGTAATCTTAATATCAACACAGAGAG 420
DB 817 TTTGACATATGATTAATTTGTTCCAAATGATCCGTAATCTTAATATCAACACAGAGAG 876
QY 421 TGGTCAGGTCTGCTACTTTAAACCAAGATGCCAAGCCGTTATTTCTATACAGATATAC 480
DB 877 TGGTCAGGTCTGCTACTTTAAACCAAGATGCCAAGCCGTTATTTCTATACAGATATAC 936
QY 481 TCAGATTAATCTGGAAGATGTGAACCGGTCTGTGTAACCAATCATTTCAACTGCTCAA 540
DB 937 TCAGATTAATCTGGAAGATGTGAACCGGTCTGTGTAACCAATCATTTCAACTGCTCAA 996
QY 541 GTAACCTTATCCAGCCGGATGACAGCTACACTTAAAGTCGATGAGTATGATCATAAA 600
DB 997 GTAACCTTATCCAGCCGGATGACAGCTACACTTAAAGTCGATGAGTATGATCATAAA 1056
QY 601 TCTGCTTTGATGGCGGAGAGGATACATTTATCAAAATATTCACCAATTTATCATGAA 660
DB 1057 TCTGCTTTGATGGCGGAGAGGATACATTTATCAAAATATTCACCAATTTATCATGAA 1116
QY 661 GGCAAGTGAATTCAGGTGATACCATTTAAGAGACCCCTACTATGTGAAGATTAAG 720
DB 1117 GGCAAGTGAATTCAGGTGATACCATTTAAGAGACCCCTACTATGTGAAGATTAAG 1176
QY 721 GGCCATAAATATCTTGTCTTTGAAGCGAATCTGAGAACAGATGTTTATCAAGCGCAT 780
DB 1177 GGCCATAAATATCTTGTCTTTGAAGCGAATCTGAGAACAGATGTTTATCAAGCGCAT 1236
QY 781 CAGTCCTTCATTAATTAAGCTTACTATGGCGGAAGTGCCTCTCTCCAGAAATGAAAA 840
DB 1237 CAGTCCTTCATTAATTAAGCTTACTATGGCGGAAGTGCCTCTCTCCAGAAATGAAAA 1296
QY 841 AATAAAGCTCTTCAAGTCTAATAAACAATTTGCTTTTAGCAATGATGATTAAGC 900
DB 1297 AATAAAGCTCTTCAAGTCTAATAAACAATTTGCTTTTAGCAATGATGATTAAGC 1356
QY 901 ATTGTGATTTGGCCGATGACTATACAGTGAAGGTGTTATGAACCATTAAGTCGATCA 960
  
```

```

DB 1357 ATTGTGAATTTGCCGATGACTATACAGTGAAGGTGTTATGAACCATTAAGTCGATCA 1416
QY 961 AACACAGTAGAGATGAGATGCGAAGCGGCCCAATATTTAATTAATGAATAATATGAT 1020
DB 1417 AACACAGTAGAGATGAGATGCGAAGCGGCCCAATATTTAATTAATGAATAATATGAT 1476
QY 1021 CTATTCACGAGATTCNAAGAGATCCCAAAATGACAGATGAGAAATTAACGACAAAGATGTT 1080
DB 1477 CTATTCACGAGATTCNAAGAGATCCCAAAATGACAGATGAGAAATTAACGACAAAGATGTT 1536
QY 1081 TATATGCTAGAGGCCCGGAGGAGACTCTTAATATGCGCCACACAAACCGATTAATGAAC 1140
DB 1537 TATATGCTAGAGGCCCGGAGGAGACTCTTAATATGCGCCACACAAACCGATTAATGAAC 1596
QY 1141 GGACTTGTATGAACATGAATCTTACCTGCTGATCTCACACACACTTACTCTCATTTGC 1200
DB 1597 GGACTTGTATGAACATGAATCTTACCTGCTGATCTCACACACACTTACTCTCATTTGC 1656
QY 1201 GGTATCCCGCACCCCTGAAGGTAAATATGTTGTTACTCACAGTATATAGCAATAGAGGC 1260
DB 1657 GGTATCCCGCACCCCTGAAGGTAAATATGTTGTTACTCACAGTATATAGCAATAGAGGC 1716
QY 1261 TTCTATCCAGAAACATCTCTTCACCTCGCGGACAAGCTTGGGCTTAATATTAAGGCTCT 1320
DB 1717 TTCTATCCAGAAACATCTCTTCACCTCGCGGACAAGCTTGGGCTTAATATTAAGGCTCT 1776
QY 1321 GACACATCTGGAGGAGAAATATGTTCCGAGACAGCAATTCCTCA 1365
DB 1777 GACACATCTGGAGGAGAAATATGTTCCGAGACAGCAATTCCTCA 1821
  
```

## RESULT 3

```

US-09-974-300-2644
; Sequence 2644; Application US/09/974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2644
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2644
  
```

Query Match 45.2%; Score 617.2; DB 10; Length 1449;  
 Best Local Similarity 67.3%; Pred. No. 4.2e-151;  
 Matches 896; Conservative 0; Mismatches 418; Indels 18; Gaps 1;

```

QY 12 GCATACACAGAGAGATATGTTTGGCCATATTACAGCGCTGACATGCTAAAAATTCC 71
DB 99 GGATTAACAAGAAACCTACAGGATTTTCTCATATCAAGACATGACATGCTGAATAATTC 158
QY 72 AGGACACAAAACAGCTCTCTCAATTTAAAGGCTCTCAATTCATGATGACATCAAAA 131
DB 159 CGAGGAGCAAAAAGCAGCAACATTTAAAGTCTCAATTCATGCGAAGAAACATCAAAA 218
QY 132 CATTCATTCGGCAAAAGGATATGATAGTCAGGCAACTTAATAGATTAGATGATGGA 191
DB 219 CATTCCTTCTGCAAAAGGATATACAAAATGAGAGCTGATCGATTTAGACGATGGA 278
QY 192 TAGCTGGCCACTGCAAAAACGCTGATGATGCTGCGCAATTAATCATGATATCATATCGT 251
  
```



1417 GTTAACTGATGAGTACAGAGCTCT--TTGAACATCAACGGCTGTAGAGATTATATAA 1473  
601 TCTGCTTTGATGGCGAGAGCTAGATTTATCAAAATATTCAGCATTTATGATGAA 660  
1474 TCAATCTTTGA---CGGTGACGAGAAAAGCTATCAAAATGACAGCATTCATGATGAA 1530  
661 GGCAAGTGATTTGAGGTGATTAACCATTTTAAGAGACCTCACTATGTGAAGATTAAG 720  
1531 GGCAACTATACAGCTCAGGGGACCAACCATACGCTGAGAGATCTCACTACGTAGAGATTA 1590  
721 GGCCATAATATCTTCTTTGAAGCGAATCTGGAACACAGATGTTATCAAGCGAT 780  
1591 GGCCACAATATCTTGAATTTGAAGCAAACTGGAATGAGATGCTTACCAAGCGCAA 1650  
781 CAGCTTTCAATATTAAGCTTACTATGCGGAGTACGCTCTCTTCCCAATGAATAA 840  
1651 GAATCTTTATTTAAACAGCATTAATGGAACAAAGCATCATCTTCTCCCAAGAAAGT 1710  
841 AATAAAGTCTTCAAGTCTTAAACAAATGCTCTTTAGCGAATGTCATTTAGC 900  
1711 CAATAAAGTCTTCAAGGATTAATAAAGCGCATGATTTAGCAAGCGCTCTCGGT 1770  
901 ATTCTTGAATTTGCGGATGACTATACAGTGAAGATTTATGAACCATTTAGTCGATCA 960  
1771 ATGATTTGAGCTTAAAGATGATTAACACTGAAAAAGATGAACCGCTGATTCATCT 1830  
961 AACAGATGATGATGAAGTGAAGCGCAATATTTAAATGAATTAATGAT 1020  
1831 AACAGATGATGAAGTGAAGCGCAATATTTAAATGAATTAATGAT 1890  
1021 CTATTCAGGATTCAGAGATTCAGAAATGAGATGATGATTAACAGCAAGATGAT 1080  
1891 CTGTTACTGATGCTCCCGGATTCAGAAATGAGATGATGATTAACAGCAAGATGAT 1950  
1081 TATATGATGAGGCGGAGGAGGATCTTAAATGCGGACCAACCGATTAATGAAT 1140  
1951 TACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2010  
1141 GCACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
2011 GCGCTTGTGTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2070  
1201 GGTATCCGAGCAGCTGAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1260  
2071 GCTTACTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2130  
1261 TTTATCCAGAGATCTCTACCTGCGGAGCAAGCTTGGGTTAATTAATTAAGGCTCT 1320  
2131 TTTATCCAGAGATCTCTACCTGCGGAGCAAGCTTGGGTTAATTAATTAAGGCTCT 2190  
1321 GACATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1343  
2191 AAAACATCTGTTGCAAGAGCAG 2213

RESULT 5  
US-09-993-292A-18  
Sequence 18, Application US/09993292A  
Patent No. US20020146430A1  
GENERAL INFORMATION:  
APPLICANT: James E. Galen  
TITLE OF INVENTION: USE OF CLY A HEMOLYSIS FOR EXCRETION OF  
FILE REFERENCE: HOFMD 007A  
CURRENT APPLICATION NUMBER: US/09/993, 292A  
PRIOR FILING DATE: 2001-11-23  
PRIOR APPLICATION NUMBER: 60/252,516  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18

LENGTH: 8908  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: pSEC84sacB vector  
US-09-993-292A-18

Query Match 42.4%; Score 578.6; DB 10; Length 8908;  
Best Local Similarity 67.2%; Pred. No. 1.4e-140;  
Matches 903; Conservative 0; Mismatches 389; Indels 51; Gaps 4;

1 ATGAACAGCGGGGACCTACAGAGAGATGATGATTTGGCCATTTATACAGCGGCTGACATG 60  
1437 AGCAACCAAAAGCATATTAAGAAACATACGAGCATTTCCATATTAACAGCGCATGATATG 1496  
61 CTAAAAATTCAGAGCAACAAACAGCTCTAATTTAAAGTCCCTCAATTTATGATCA 120  
1497 CTGCAAAATCCCTGAACCAAAATGAATAATGAATGAATGAATGAATGAATGAATGAAT 1556  
121 GCAATCAAAACATTTGATTCGCAAAAGGATGATGATGATGATGATGATGATGATGAT 180  
1557 ACAATTAATAATATCTCTCTTCCAAAGC-----CCTG 1589  
181 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
1590 GACGTTTGGAGAGCTGCGCATTTACAAACAGCTGACGCTGACGCTGACGCTGACGCTGAC 1649  
241 TATCATATCTCTCCGCTTTAGCAGGTACCCCAAAACAGTATGATATCTCCATTCAT 300  
1650 TACACATCT 1709  
301 TTAATCT 360  
1710 ATGTTCT 1769  
361 TTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
1770 TTTAAAGACAGCAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1829  
421 TGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
1830 TGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1889  
481 TCAGATATCTCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1931  
1890 TCCGTTAAACATTA-----CGGCAACCAACAGCTACAGCTACAGCTACAGCTACAG 1991  
541 GTAACTTATCCAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
1932 GTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1988  
601 TCTGCTTTGATGGCGGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
1989 TCAATCTTTGA---CGGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2045  
721 GGCCATAATATCTTCTTTGAAGCGAATCTGGAACACAGATGTTATCAAGCGAT 780  
2106 GGCCACAATATCTTGAATTTGAAGCAAACTGGAATGAGATGCTTACCAAGCGCAA 2165  
781 CAGCTTTCAATATTAAGCTTACTATGCGGAGTACGCTCTCTTCCCAATGAATAA 840  
2166 GAATCTTTATTTAAACAGCATTAATGGAACAAAGCATCATCTTCTCCCAAGAAAGT 2225  
841 AATAAAGTCTTCAAGTCTTAAACAAATGCTCTTTAGCGAATGTCATTTAGC 900  
2226 CAATAAAGTCTTCAAGGATTAATAAAGCGCATGATTTAGCAAGCGGCTCTCGGT 2285  
901 ATTGTTGAATTTGCGGATGACTATACAGTGAAGATTTATGAACCATTTAGTCGATCA 960

Db	2286	ATGATTTAGCCTAAACAGATGATTTACACACTGAAAAAAGATGAAACCGCTGATTTGCATCT	2345
Qy	961	AACACAGTAGCAGATGAAGTCGAAACGGCCCAATATATTTAAATGAATATAAATGCTAT	1020
Db	2346	AACACAGTAAACAGATGAAATTTGAACCGGCGCAACCTTTTAAAAAGAACGGCAAAATGGTTAC	2405
Qy	1021	CTATTCCAGGTTTCAAGAGATGCCAAAATGACGAGTATGGAATTTAAACGACAAAGATGTT	1080
Db	2406	CTGTTTCACTGCTCTCCCGGATCAAAAATGACGATTCAGCGCATTTAGCTCTAACGATTT	2465
Qy	1081	TATATGCTAGGGCCCGGAGGCGACTCCCTTCTTAATATGCGCCACAAACCGATTAATGAATCT	1140
Db	2466	TACATGCTGGTTATGTTTTCTTAATTTCTTTAACTGGCCCATATCAAGCGCGTGAACAAACT	2525
Qy	1141	GGACTTGTATTGAACATGATCTTGACCCCTCTATCTCACACACACTTACTCTCATTCG	1200
Db	2526	GGCCCTTGTTTTAAAAATGATCTTGATATCTCAAGCATGTAACTTACTTACTTCACACTTC	2585
Qy	1201	GGTATCCCGCACCCCTGAAGGTAATTAATGTGTACTCACAACTTTATGACGAATAGAGGC	1260
Db	2586	GCTGTACTCTCAAGCGAAAGAAACAAATGTCTGTATTACAGCTTATATGACAAACAGAGCA	2645
Qy	1261	TTCTATCCAGAACATCACTCTCACCTGCGGACAAGCTTGGGGTTTAATATTTAAAGGCTCT	1320
Db	2646	TTCTAAGCAGACAAACATCAACGTTTGGCGCAAGCTCTCTGTAACATCAAGAGCAAG	2705
Qy	1321	GACACATCTGGAGGAGAAAAATAG	1343
Db	2706	AAACATCTGTGTCAAAAGACAG	2728

```

RESULT 6
US-10-055-794-1
: Sequence 1, Application US/10055794
: Publication NO. US20030059900A1
: GENERAL INFORMATION:
: APPLICANT: Farmer, Alan Andrew
: TITLE OF INVENTION: SEQUENCE SPECIFIC RECOMBINASE-BASED
: TITLE OF INVENTION: METHODS FOR PRODUCING INTRON CONTAINING VECTORS AND
: FILE REFERENCE: CLON-069
: CURRENT APPLICATION NUMBER: US/10/055,794
: CURRENT FILING DATE: 2002-01-17
: PRIOR APPLICATION NUMBER: 60/263,358
: PRIOR FILING DATE: 2001-01-18
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 4938
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic Vector
US-10-055-794-1

```

Query Match	42.38;	Score 577;	DB 9;	Length 4938;
Best Local Similarity	67.28;	Pred. No. 2.6e-140;		
Matches 902; Conservative	0;	Mismatches 390;	Indels 51;	Gaps 4;

QY 1 ATGACACGCGGGGACCTACAGGAAGACTATG6TTTGGCCATPNTTACAGCGGCTGCACATG 60  
 Db 1602 ACGAACCMAAAGCCATATTAAGGAAACATAGCGCATTTCCCATATTTACAGCGCATGATNTG 1661  
 QY 61 CTAAAAATTCAGGACACAACAAACAGTCTCAATTTAAAGTGCCCTCAATTCAATGATCA 120  
 Db 1662 CTGCAAAATCCCTGAACACAGCAAAAAAATGAATAATATCAAGTTCCGAGTGTGATGTCGHC 1721  
 QY 121 GCATCAAAAAATGATGTCGGCAAAAGGATATGATTAAGTCAGGCACTTAATGATTTA 180  
 Db 1722 ACAATTAATAATATCTCTTCGCAAAAG-----CCTG 1754  
 QY 181 GATGATG6GATAGCTGGCCACTGCAAAAGCCTGATGTTACTGGCGCAATATTCATGCA 240

Db	175	GACGTTTGGGACACCTGGCCATTTCACAAACCGCTGACGGCACTGTGGCAAACTATCAGCG	1814
Oy	241	TATCAATCGTCTCCGCTTTAGCAGGTGACCCAAAACAGTGTGATCTCCATTCAT	300
Db	1815	TACCACATCGCTTTGCATTATGCCGGAGATCTCTAAAAATGGCGATGACATCGATTAC	1874
Oy	301	TTATTCTATCAAAAAGTCGGTATACATTCGATTACAGCTGTGAAAAATGCTGTGAAGGTA	360
Db	1875	ATGTTCATATCAAAAAGTCGGGAAACCTCTAATGACAGCTGAAAAACGCTGGCGCGTC	1934
Oy	361	TTTGAAGATATGATTAATTTGTTCCAAAATGATCCGATCTTAAATATCAACACAGAG	420
Db	1935	TTTAAAGACAGCGACAAATTCGATGCAATGATTTCTTCTTAAAGACCAACACAGAA	1994
Oy	421	TGTCAGGTTCTGCTACTTTAACCAGAAATGGCCAGTCCGTTTATTTCTATACAGATTAC	480
Db	1995	TGTCAGGTTCTGCGACCATTTTACATCTGACGGAAAAATCCGTTTATTTCTACAGATTTC	2054
Oy	481	TCAGGTAATCTGAAAGATGGTGGAAACGGTGCTGTGTAAACCAATCATTTCACTGCTCAA	540
Db	2055	TCGCGTAAACATTT-----CGGCAACCAACATCGACAACTGCACAA	2096
Oy	541	GTAACATTAATCCGACCGGATGCAGCTACACTTAAATCGATGGATATCGATCATAA	600
Db	2097	GTTAAAGTATGAGATCAGACAGACGCTCT---TTGAACATCAACGGTGTGAGAGATTATAA	2153
Oy	601	TCGTCTTTGATGGCGGAGCGGTACAGTTTATCMAAATATTCAGCAATTTATGATGAA	660
Db	2154	TCATCTTTGA---CGGTGACGGAACAAACGATCAAAATGTACAGCATTCATGATGAA	2210
Oy	661	GGCAAGTGTGATTTCAAGTGTATACCATTACTTTAAGACACCTCTATGTTGAAGATAG	720
Db	2211	GGCAACTACACCTCAGGCGACACCAATACGCTGAGAGATCTCTCCTAGTGAAGATAAA	2270
Oy	721	GGCCATAATATCTTGTCTTTGAAGCGCAATCTGGAACACAGATGTGTTATCAAGCGAT	780
Db	2271	GGCCACAAATCTTAGTATTTGAAGCAACACTGGAATGAAGATGGCTACCAAGCGCAA	2330
Oy	781	CAGTCTTTCAATTAATAAGCTTTACTATGGCGGAAGTAGCTCTTTCTCCAGAAATGAAAA	840
Db	2331	GATCTTTAATTAACAAGCAATCTATGGCAAAACATCATTTCTTCCGTCAAGAAAGT	2390
Oy	841	AATTAATGCTTCAAGTCTTAAAAAACAAATTCCTTTTACGGAATGCTGATTTAGCG	900
Db	2391	CAAAAACCTTCGCAAGCGATAAAAAACGCGCTAGTTAGCAAAACGCGCTCTCGGT	2450
Oy	901	ATGTGTGAATGGCGGATGACTATACAGTGAAGTGTATGAACCATTAAGTGTGCATCA	960
Db	2451	ATGATTTGCTTAACAGATTTACACCTGTAAAAAGTGAATGAACCGCTGATTTGATCT	2510
Oy	961	AACACAGTAGCATGAAGTGAACGCGCCATATTTTAAATGAATTAATATGCTAT	1020
Db	2511	AACACAGTAACAGATGAATTTGAACGCGCGCAAGCTTTTAAATGAACGGCAATGGTAC	2570
Oy	1021	CTATTACAGATTCAGAGATCCAAAATGACGATGAGATTTAACGACAAAGATGT	1080
Db	2571	CTGTTCACGACTCCCGGATCAAAAATGACGATTAAGCGCATTTAGCTCTAACGATTT	2630
Oy	1081	TATATGCTAGGCGCCGAGGCGACTCCTTAATATGGCCACACAAACCGGATTAATGAAGT	1140
Db	2631	TACATGCTGGTATATGTTTCTAATTTCTTAACTGGCCCATTAACAGCGCTGAACAAAACT	2690
Oy	1141	GGACTGTGATTAAGACATGAATCTTGACCCTCTGATCTACACACACTTATCTCATTTGC	1200
Db	2751	GCTGACTCTCAAGGAAGAAACAATGTGCTGATTAACAAGCTATATGACAAACAGAGA	2810
Oy	1201	GGATATCCCGCACCTGGAAGGTAAATATGTGTACTCACAAAGTTATATGAGAAATAGAGC	1260
Db	2811	TTCTATCCAGACATCACTCTCACCTCGGGACACAGCTTGCGGTTAATATTAAGGCTCT	1320
Oy	1261	TTCTATCCAGACATCACTCTCACCTCGGGACACAGCTTGCGGTTAATATTAAGGCTCT	1320
Db	2811	TTCTATCCAGACATCACTCTCACCTCGGGACACAGCTTGCGGTTAATATTAAGGCTCT	2870



QY 1321 GACACATCTGAGAGAAAAATAG 1343  
 Db 2871 AAAACATCTGTGTCAAAAGACAG 2893

RESULT 7

US-10-055-794-3  
 ; Sequence 3, Application US/10055794  
 ; Publication No. US20030059900A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Farmer, Alan Andrew  
 ; TITLE OF INVENTION: SEQUENCE SPECIFIC RECOMBINASE-BASED  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING INTRON CONTAINING VECTORS AND  
 ; FILE OF INVENTION: COMPOSITIONS FOR USE IN PRACTICING THE SAME  
 ; FILE REFERENCE: C10N-069  
 ; CURRENT APPLICATION NUMBER: US/10/055,794  
 ; CURRENT FILING DATE: 2002-01-17  
 ; PRIOR APPLICATION NUMBER: 60/263,358  
 ; PRIOR FILING DATE: 2001-01-18  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 3  
 ; LENGTH: 6525  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic Vector  
 US-10-055-794-3

Query Match 42.3%; Score 577; DB 9; Length 6525;  
 Best Local Similarity 67.2%; Pred. No. 3,1e-140;  
 Matches 902; Conservative 0; Mismatches 390; Indels 51; Gaps 4;

QY 1 ATGACAGGGGGGATACAGAGACATGCTTTTCCCATATACAGCCGCTGACATG 60  
 Db 3189 AGCAACCAAAAGCCATATTAAGAAACATAGGCAATTTCCCATATTAAGCCATATGAG 3248  
 QY 61 CTAAATATTCAGAGACAAACAAAGCTCCATTAATTAAGTGCCTCAATTCATGATCA 120  
 Db 3249 CTGCAATCCCTGAGACAGCAAAATGAAATATCAAGTTCGAGTTCATTCGCTGC 3308  
 QY 121 GCAATCAAAACATTTGATGCGCAAAAGGATGATAGTCAAGCACTTAATAGATTAA 180  
 Db 3309 ACAATTAATAATATCTCTCTCAAAAGG-----CCCG 3341  
 QY 181 GATGATGAGGATGAGTGGCCACTGCAAAACGCTGATGACTGCGCAATATATCATGA 240  
 Db 3342 GACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCAGCTGCGCAAACTATCAGCGC 3401  
 QY 241 TATCAGATCGTCTCGCTTTTACAGGTGACCCAAAACAGATGATGATCCACTTCAT 300  
 Db 3402 TACCAATCGTCTTTCGATTTAGCCGAGATCTTAAATAATGCGGATGACATGATTTAG 3461  
 QY 301 TTATTTATCAAAAAGTGGTGATATCATGATGATGACAGTGAATAATGCTGAAGAGTA 360  
 Db 3462 ATGTTCTATCAAAAAGTGGCGCAAACTCTTATTTGACAGCTGGAATAAGCTGGCGGCTC 3521  
 QY 361 TTTGAAGATGATGATTAATTTGTTCGAATGATCCGATCTTAATATCAACACAGAG 420  
 Db 3522 TTTTAAAGACAGGAGCAAAATGATGATTAATTTATCTTAAAGACCAACACAAAGAA 3581  
 QY 421 TGTGAGGTCTGCTACTTTTAAACAAAGATGCGCAAGTCCGTTTATTTATACAGATTAC 480  
 Db 3582 TGTGAGGTCTGCTACTTTTAAACAAAGATGCGCAAGTCCGTTTATTTATACAGATTAC 480  
 QY 481 TCAGGTAACTCTGAAGATGTTGGAACGCTGCTGTTAATCAATATTTCAACTGCTCA 540  
 Db 3642 TCCGTTAAACATTTA-----CGGCAAAACAAACACTGACACACTGACACAA 3683  
 QY 541 GTAACCTTATCCAGCCGATGACATCACTTAAGTCAAGTGAATGATATATATATAA 600  
 Db 3684 GTTAAAGTATGATGATGACAGAGCTCT--TTGAACATCAAGGTTAGAGATTATATAA 3740

QY 601 TCTGCTTTTATGCGGACGCTACAGTGTATCAAAAATATTCAGCAATTTATGAGTAA 660  
 Db 3741 TCAATCTTTGA---CGGTACGCGGAAAAAGCTATCAAAAATGTACAGAGTTCATGATA 3797  
 QY 661 GGCAGTGGATTTTCAGGTGATTAACCATCTTTAAGAGACCCCTACTATGTTGAAGATA 720  
 Db 3798 GGCAGTGGATTTTCAGGTGATTAACCATCTTTAAGAGACCCCTACTATGTTGAAGATA 720  
 QY 721 GGCAGTGGATTTTCAGGTGATTAACCATCTTTAAGAGACCCCTACTATGTTGAAGATA 780  
 Db 3858 GGCAGTGGATTTTCAGGTGATTAACCATCTTTAAGAGACCCCTACTATGTTGAAGATA 780  
 QY 781 CAGTCTTTAATTAATTAACCTTACTATGCGGAGAGAGAGCTCTCTCAGATGATAA 840  
 Db 3918 GATCTTTTATTAACCAACATCTATGTCGCAAAAGACATCTTCTCCGTTAAGAAAGT 3977  
 QY 841 AATTAACGCTTCAAAAGTCTTAAACCAAAATGCTTTTAAAGCAATGCTCATTAAGC 900  
 Db 3978 CAATAACCTCTGCAAAAGGATTAATAAAGCAGCGCTGAGTTGCAAAAGCGGCTCGGT 4037  
 QY 901 ATTTGATGATTTGGCGGATGATATACAGTGAAGAGTGTATGAACCATTAAGCTATCA 960  
 Db 4038 ATGATTTGATGATTTGGCGGATGATATACAGTGAAGAGTGTATGAACCATTAAGCTATCA 960  
 QY 961 AACACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 Db 4098 AACACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 QY 1021 CTATTCAGGATTTCAAGAGATCCAAATATGACAGATGATGATGATGATGATGATGATGAT 1080  
 Db 4158 GTTTCAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 1081 TATATCTATGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140  
 Db 4218 TACATGCTGTTGTTATGTTTCTAATTTCTTAATGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140  
 QY 1141 GACATGTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 Db 4278 GGCCTTTGTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 QY 1201 GGTATCCGACCCCTGAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
 Db 4338 GGTATCCGACCCCTGAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
 QY 1261 TTTCTATCCAGATCACTCTCACTGCGGCAAAAGCTTTGGGTTAATTAAGAGTCT 1320  
 Db 4398 TTTCTATCCAGATCACTCTCACTGCGGCAAAAGCTTTGGGTTAATTAAGAGTCT 1320  
 QY 1321 GACACATCTGGAGGAAAAATAG 1343  
 Db 4458 AAAACATCTGTGTCAAAAGACAG 4480

RESULT 8

US-09-849-866-1/C  
 ; Sequence 1, Application US/09849866  
 ; Patent No. US2002006833A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ilya Chumakov  
 ; APPLICANT: Hiroaki Tanaka  
 ; TITLE OF INVENTION: High Throughput DNA Sequencing Vector  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knodbe, Martens, Olson & Bear, LLP  
 ; STREET: 550 West C Street, Suite 1200  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 92101  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: win95
SOFTWARE: word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/849,866
FILING DATE: 04-MAY-2001
ATTORNEY/AGENT INFORMATION:
NAME: Daniel Hart
REGISTRATION NUMBER: 40,637
REFERENCE/DOCKET NUMBER: GENSET.15CDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10317 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: CIRCULAR
MOLECULE TYPE: synthetic DNA
ORIGINAL SOURCE:
ORGANISM: cloning vector pGendel
FEATURE:
NAME/KEY: pGendel
LOCATION: 1..10317
FEATURE:
NAME/KEY: Homology with X06404 comp1 (411..1668)
LOCATION: 9..1266
IDENTIFICATION METHOD: blastn against X06404
FEATURE:
NAME/KEY: Kanamycin resistance gene CDS
LOCATION: 142..957
IDENTIFICATION METHOD: By homology to X06404
FEATURE:
NAME/KEY: Tn1000's right end
LOCATION: complement 1332..1371
IDENTIFICATION METHOD: blastn against X06200)
FEATURE:
NAME/KEY: Homology with U46017 (1-472)
LOCATION: 1423..1894
IDENTIFICATION METHOD: blastn against U46017
FEATURE:
NAME/KEY: single stranded DNA replication origin
LOCATION: 1423..1894
IDENTIFICATION METHOD: By homology to U46017
OTHER INFORMATION: mutation T -> C 1658
FEATURE:
NAME/KEY: Homology with U51113 (2382..6997)
LOCATION: 1896..6544
IDENTIFICATION METHOD: blastn against U51113
FEATURE:
NAME/KEY: OriS
LOCATION: 1972..2188
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: repELR
LOCATION: 2897..2918
OTHER INFORMATION: Described in seqid 16
FEATURE:
NAME/KEY: RepE
LOCATION: 2903..3034
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: T3
LOCATION: 3043..3059
OTHER INFORMATION: Described in seqid 17
FEATURE:
NAME/KEY: LRT3RA
LOCATION: complement 3045..3069
OTHER INFORMATION: Described in seqid 15
FEATURE:
NAME/KEY: IncC
LOCATION: 3070..3320
IDENTIFICATION METHOD: By homology to U51113

```

```

OTHER INFORMATION: insertion 33 bases 3038..3071
FEATURE:
NAME/KEY: ParA
LOCATION: 3655..4821
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: mutation G -> A 3878
FEATURE:
NAME/KEY: ParB
LOCATION: 4821..5792
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: ParC
LOCATION: 5865..6382
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: Homology with J01688 (complement 175..819)
LOCATION: 6574..7218
IDENTIFICATION METHOD: blastn against J01688
OTHER INFORMATION: mutation A -> G 7096
FEATURE:
NAME/KEY: CDS streptomycin sensitivity gene
LOCATION: complement 6716..7090
IDENTIFICATION METHOD: By homology to J01688
OTHER INFORMATION: mutation A -> G 6728
mutation G -> C 6821
mutation C -> T 6866
mutation T -> C 7013
mutation T -> A 7058
FEATURE:
NAME/KEY: rpsLr
LOCATION: 7155..7174
OTHER INFORMATION: Described in seqid 12
FEATURE:
NAME/KEY: SP6
LOCATION: 7230..7248
OTHER INFORMATION: Described in seqid 13
FEATURE:
NAME/KEY: Tn1000's left end
LOCATION: 7252..7291
IDENTIFICATION METHOD: blast (X60200)
FEATURE:
NAME/KEY: Homology with X02730 (complement 37..1959)
LOCATION: 7305..9227
IDENTIFICATION METHOD: blastn against X02730
FEATURE:
NAME/KEY: CDS levansucrase gene
LOCATION: complement 7379..8800
IDENTIFICATION METHOD: By homology to X02730
OTHER INFORMATION: mutation T -> C 7466
mutation A -> G 7739
mutation T -> C (Asn -> Asp) 8347
mutation T -> C 8600
mutation G -> A (Ala -> Val) 8772
FEATURE:
NAME/KEY: SLR3
LOCATION: 8711..8731
OTHER INFORMATION: Described in seqid 14
FEATURE:
NAME/KEY: Homology with J01636 (complement 1158..1465)
LOCATION: 9298..9623
IDENTIFICATION METHOD: blastn against J01636
FEATURE:
NAME/KEY: CDS alpha peptide beta-galactosidase
LOCATION: complement 9276..9497
IDENTIFICATION METHOD: By homology to J01636
FEATURE:
NAME/KEY: primer HE1
LOCATION: complement 9465..9479
FEATURE:
NAME/KEY: primer HE2
LOCATION: 9461..9475
FEATURE:
NAME/KEY: primer LacLR52Avr

```

```

;
; LOCATION: complement 9603..9630
;
; FEATURE:
;   NAME/KEY: primer LAGE2M1u
;   LOCATION: 9289..9314
;
; FEATURE:
;   NAME/KEY: Homology with M77789 (1889..2576)
;   LOCATION: 9629..10315
;   IDENTIFICATION METHOD: blastn against M77789
;
; FEATURE:
;   NAME/KEY: high copy-number double-stranded DNA replication origin
;   LOCATION: complement 9629..10315
;   IDENTIFICATION METHOD: By homology to M77789
;
; OTHER INFORMATION: mutation C -> T 9803
;
; site PmlI 10029 - 10043
;
; CLONING SITES 10031 - 10041
;
; FEATURE:
;   NAME/KEY: oriLRd
;   LOCATION: 9856..9881
;   OTHER INFORMATION: Described in seqid 8
;
; FEATURE:
;   NAME/KEY: OSI
;   LOCATION: 10009..10026
;   OTHER INFORMATION: Described in seqid 10
;
; FEATURE:
;   NAME/KEY: ORI
;   LOCATION: complement 10046..10062
;   OTHER INFORMATION: Described in seqid 11
;
; FEATURE:
;   NAME/KEY: oriLRt
;   LOCATION: complement 10182..10202
;   OTHER INFORMATION: Described in seqid 9
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
;
; US-09-849-866-1
;
; Query Match      42.3%; Score 577; DB 10; Length 10317;
; Best Local Similarity 67.2%; Pred. No. 3.9e-140;
; Matches 902; Conservative 0; Mismatches 390; Indels 51; Gaps 4;
;
;
; 1 ATGAACAGCGGGGACACTCAAGAGAGACTATGAGTTTGGCCATATTACAGCGCGTACATG 60
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8707 ACGAACCAAAAGCCATATAGGAAGCAATAGCGCATTTCCATATTTACAC3CCATGATATG 8648
;
; 61 CTAAAAATTCAGAGACAACAAGAGTCTCAATTTAAAGTCCCTCAATTCATCATCA 120
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8647 CTGCAAAATCCCTGAACACAAAAAATTAATTAAGTTCTCGAGTTGATTCGTC 8588
;
; 121 GCATCAAAAACATTTGATTCGGCAAAAGGATATGATTAAGTCAGCGCACTAATAGATT 180
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8587 ACAATTTAAAAATATCTCTTCTGCAAAAAG-----CCTG 8555
;
; 181 GATGTATGGATAGTCTGCGCACTGCAAAAGCTGATGTGCTACGCGCAATTATCATGGA 240
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8554 GACGTTTGAGAGAGTGGCCATTACAAAAGCTGACGCGCTGCAACATATCAGCGC 8495
;
; 241 TATCAGATCGTCTCCGCTTATAGAGGTGACCCAAAAAAGAGATGATATCTCACTTCAT 300
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8494 TACCACATCGTCTTTCGATTTAGCGGAGATCTCTAAATTTGGGATGACACATTCATTAC 8435
;
; 301 TTATCTATCAAAAAGTGGTGATACATGATTCAGACCTGGAAAAATGCTGGAGAGTA 360
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8434 ATGTTCTATCAAAAAGTGGGGAACCTTCTATTCACACTGGAAAAAGCTGGCGCGTC 8375
;
; 361 TTGAGATATGATTAATTTGTTCCAAATGATCGTATCTTAATATATCAACACAGAG 420
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8374 TTTAAAGACACGACGACATTTGATGACGATATTTCTATCTTAAGACCAACAAAG 8315
;
; 421 TGTGACAGTTCTGCTACTTTAACCAGAGATGCGCAAGTCCCTTATTTATATACAGATTAC 480
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8314 TGTGACAGTTCTGACCATTTTACATCTGACGGAATAAATCCGTTTATTTACATCATTTTC 8255
;
; 481 TCAGTAACTCTGAAGATGTGGAACCGGTCTGCTAGTAACCAATTCATTTCAATGCTTCA 540
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

;
; Db 8254 TCCGCTAAACATTA-----CGGCAAAACAAGACTGACAACTGCACAA 8213
;
; 541 GTAACCTTATCCCGGATGACAGCTACACTTAAGTCGATGATGATTCGATCATTA 600
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8212 GTTACGATACGATCAGACAGCTCT--TTGACATCAACGGGTAGAGATTATTA 8156
;
; 601 TCTGCTTTTGAATGGCGAGAGGTACACTTATCAAAATATTCACAAATTTATGAGTA 660
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8155 TCAATCTTTGA--CGGTACGGAATAACGATATCAAAATGATACAGCACTTCATCATGA 8099
;
; 661 GGCAAGTGATTTACGATGATACCATATCTTTAAGAGACCTCACTATGTTGAGATAG 720
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8098 GGCAACTACAGTCTAGGCGACAAACCATATCGGTAGAGATCTCTACGTAGAGATAAA 8039
;
; 721 GGCCATTAATATCTGCTTTGAACGGAATCTGGAACACAGATGTTATCAAGCGCAT 780
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 8038 GGCCACAAATCTTATGATTTTGAACAAACATGGAACCTGAAGATGGCTACCAAGCGCA 7979
;
; 781 CAGTCTTTCAATTAATAAGCTTACTATGCGGAAGTGAAGCTCTTCTCCAGAAATGA 840
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7978 GAATCTTATTTAACAAGACATACATATGCAAAACACATCATCTTCTCCGTCAAGAA 7919
;
; 841 AATAAAGCTCTTCAAGTCTTAATAAACAATGCTTTTATAGCAATGTCATTTAGCC 900
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7918 CAAAACCTCTGCAAAAGCGATAAAAACGACGCTGATTAAGCAAAAGCGCGCTCTCGGT 7859
;
; 901 ATTGTTGAATGGCCGATGACTATACAGTGAAGAAATGTTATGAACCATTTAGTGCATCA 960
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7858 ATGATTAGCTTAACAGATGATTTACACACTGAAAAAGATGATGAACCGCTGATGCATCT 7799
;
; 961 AACACAGTAGACATGATGATGCAAGCGGCCCAATATATTTAAATTAATAATAGTAT 1020
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7798 AACACAGTAGACATGATGATGCAAGCGGCCCAATATATTTAAATTAATAAGCGCAATAGTAT 7739
;
; 1021 CTATTCACGATTTCAAGAGATCCAAATAGCAGTATGATGAATTAACGACAAAGATTT 1080
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7738 CTGTTCACTGATCTCCCGGATCAAAATGATGATGATGATGATGATGATGATGAT 7679
;
; 1081 TATATGCTAAGGCGCCGAGGCGACCTCTTAATATGCCCCACACAAACCGATTAATGA 1140
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7678 TACATGCTTGATGATATGTTCTTAATCTTAACTGCGCCATCAACGCGCTGAAACAACT 7619
;
; 1141 GGACTTGTATGAACATGAATCTTGACCTGCTGATCTCACACACATCTTACTCTATGC 1200
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7618 GGCCTTGTGTTAAATATGATCTTGTATCTCAAGATGATGATGATGATGATGATGAT 7559
;
; 1201 GGTATCCGACCCCTGAAGATTAATATGTTGTTACTCAAGATTAATAGCAATAGAGGC 1260
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7558 GCTGTACCTCAAGCGAAGAAACAATGTCGATTAACAGCTATATGACAAACGAGGA 7499
;
; 1261 TTCTATCGAAGAACATCACTCTCAGCTGCGGACAGCTTGGGTTAATATTAAGGCTCT 1320
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7498 TTCTAGCGAGCAAAACATCAACAGTTTCCGCGGACCTTCTGCTGAACATCAAGCGCAAG 7439
;
; 1321 GACACATCTGGAGAGAAATAG 1343
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 7438 AAAACATCTGTTGCAAAAGACAG 7416
;
;
; RESULT 9
; US-09-864-761-21746
; Sequence 21746, Application US/09864761
; Patent No. US20020048763A1
;
; GENERAL INFORMATION:
;   APPLICANT: Penn, Sharon G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   APPLICANT: Chen, Wensheng
;
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmiga-X-1
; CURRENT APPLICATION NUMBER: US/09/864.761
; CURRENT FILING DATE: 2001-05-23

```

```

NAME/KEY: unsure
LOCATION: 362
NAME/KEY: unsure
LOCATION: 363
NAME/KEY: unsure
LOCATION: 364
NAME/KEY: unsure
LOCATION: 365
NAME/KEY: unsure
LOCATION: 366
NAME/KEY: unsure
LOCATION: 367
NAME/KEY: unsure
LOCATION: 368
US-09-864-761-21746

Query Match
Best Local Similarly 64.0% Score 135.6; DB 10; Length 368;
Matches 235; Conservative 0; Mismatches 105; Indels 27; Gaps 1;

16 TACAAAGAACACATATGTTTGGCCATATTACACGGCGCTGACATGGTAAATAATCCAGCA 75
|||||
12 TACAAAGAAACGTAACGCGCTCTCTCATATTACACGCCATATATGCTGCGATCCCTAA 71
|||||

76 CAACAAAACAGCTCCTCAATTTAAAGTGCCCTCAATTCATGCATCAGCAATCAAAAACATT 135
|||||
72 CACAGACAAAACGAAAATATCCAAAGTSCCTCAATTCGATCAATCAACGATTTAAATAATTT 131
|||||

136 GATTGGCAAAAGGGTATGATTAAGTCAGCAACTTAATTAATTAATGATGATGATAGC 195
|||||
132 GAGCTGTGCAAAAGGACTT-----GATGTGTGGGACAGC 164
|||||

196 TGGCCACTGCAAAACGCTGATGTGCTACGCGCAAAATTATCATGATATCATCTCTCC 255
|||||
165 TGGCCGCTGCAAAACGCTGACGAGACAGTAGTGAATATACAACGGCTATCACGTTGCTTT 224
|||||

256 GCTTTAGCAGGTGACCCAAAACAGTGAATGATGATCTCACTTATTTATTTATCAAAA 315
|||||
225 GCTCTTGGCGGGAAGCCCGGAAAGCGCTGATGACACACATCAATCAATGTTTATCAAAAG 284
|||||

316 GTGCGTATCATCTGATTCGATTCGACACTGGAATAATGCTGGAAGATATTTGAAGATATGAT 375
|||||
285 GTGCGGCAACAACCTCAATGACACAGCTGGAATAACGCGGCGGTCTTTNTTAAGACAGCGAT 344
|||||

376 AAATTTG 382
|||||
345 AAGTTTG 351
|||||

RESULT 10
US-09-864-761-5018
; Sequence 5018, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27

```



OY	254	CCGCTTAGCAGGTGGACCCAAAAAAGTATGTAFACTCCACTTCATTATTTATCAAA	313
Db	854	TGCGCAATGATGGGAATTCCAAACCAAA---ATGATAATCATATCTATCTTTATATA	910
OY	314	AAGTCGGGTACATGCATGTGACAGCTGGAAAATGCGGAAAGATTTTGTAAGATATGG	373
Db	911	AGTATGGTGAATATGAATTTAAGTCATTGGAGAAGATGAGTCCTCAATTTTTGGCTATATTT	970
OY	374	ATAAATTTGTTCCAATATGATCCGTATCTTAAATATCAACAACAGAGTGGTCAGGTTCTG	433
Db	971	CTACCCGGCGTT-----TCCAGAAATGGTCAGAGATCAG	1003
OY	434	CTACTTTAACCAAGATGGCCAGCTCGTTTATTTCTATACAGATTACTCAGGTAACTCTG	493
Db	1004	CTGTTTGGACAGGATTAACCTCATCTCCAAATTTTATATACA-----AGGGTAG	1051
OY	494	AAGATGGTGAACCGGTGCTGCTGTAACCAAACTTTTCAACGTCAGSTAACTTAATCCC	553
Db	1052	ACAGCTGTGATACCAATATCCAAATATCAAAAATTTGCTAGGCGTACTTTATTTAACTG	1111
OY	554	AGCCGATGTCAGCTACACTTAAAGTCGATGAGATCTGATCATATAATCTGCTTTGATG	613
Db	1112	ATATATATGGAATATGATACACTGCCCTCAGGTACGAATGACTATA-----TTGATATTGG	1165
OY	614	GCGGAGACGGTACAGTTTATCAAAATATTCGAATTTATGATGAAGGCAAGTGATTT	673
Db	1166	AAGGTGATGGCTATTACTACCAAACTTATGATCATGGAAGGAACTACTACAAAG--TG	1222
OY	674	CAGGTGATAACCATTAATTTAAGACACCCCTCACTATGTTGAAGATTAAGGCCATTAATATC	733
Db	1223	CCGATTAATATTCGATGCGTGATCTCATGTAATTTGAAGATGTAATGCTGATCGTATCC	1282
OY	734	TTGCTTTGAAGCAATACTGGAACAACAGATGGTATCAAGGGCATCAGTCTTTCATA	793
Db	1283	TTGTTTTTGAAGCAAGTATCGS---TTTGGAAATTTATCAAGGGAGGACCAAAATTTATA	1339
OY	794	ATAAAGCTTACTATGGCGGAAGTACGTCCTTCTCCAGATGAaaaaaaATTAACCTGCTTC	853
Db	1340	ACTGGTTAAATTAATGCGGAGATGACGAGATTTAATATCAAGGCTTTATTTGAATTCCTTT	1399
OY	854	AAAGTCCTAAAAAACCAAT---GCTTCCTTAAGGAATGGTGCATTTAGGCTTTGTGAT	910
Db	1400	CCAATGATGATTTAAGTGTGGGCAACTTGGCTTAATGACGCTATGCTATCCCAAC	1459
OY	911	TGGCGGATGACTA-----TACAGTGAAGATGTTATGAACCATTAATGCGCATCAA	961
Db	1460	TAAATTAAGACGAAGAAGATCCTTAAGGTGAGAGATTAATCTCACATTAATTTCTGAC	1519
OY	962	ACACAGTAGCAGATGAAGTCGACGCGCCCAATTAATTTAAATGAATTAATTAATGTTATC	1021
Db	1520	CAATGTGAAGCGATGAATTAAGGACCAAAATGATTAATTAATTAATTAATTAATTAAT	1579
OY	1022	TATTCACGGATTCGAAGAGATCCAAATATGACGATGAGTAATTAACGACAAAGATGTTT	1081
Db	1580	TATTTGCGGTACCCGTTTAATGAGAGAACTAATGATGATCTTGGATGAATCTAATT	1639
OY	1082	AT-----ATGGTAGGGCGGAGGAGCTCCCTTAATATGGCCCCAC	1120
Db	1640	ATGCGCGTTGGTGAATATGTTCCAAATGTCGATATGTTGCTGATATGCTTAATCTGGATTT	1699
OY	1121	ACAACCGGATTAATATGAATGAGTCTGTTATGAACATGAATCTTACCCCTGCTATATCA	1180
Db	1700	ATAAGCCATTAAATGATTTCTGAGTATGCTTGACTGCTTGCTTCGCAAACTGGCGGA	1755
OY	1181	CACACACTTACTCTCATTTGCGGTATCCCGCACCTG---AAGGTAAATATGTGTAATCA	1237
Db	1760	CAGCAACTTATTCATTAATGCTGTCCCGCTGGCGGAAAGATGACCAATATTAATAGTTA	1819
OY	1238	CAAGTTATATGACGAATAGA	1257
Db	1820	CTTCATATATGACHTAATAGA	1839

RESULT 14  
 US-09-995-587A-3  
 ; Sequence 3, Application us/09995587A  
 ; Patent No. US2002127681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN HIJUM, SACHA ADRIANUS FOKKE TACO  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDIJNA HENDRIKA  
 ; APPLICANT: DIJKHUIZEN, LUBBERT  
 ; APPLICANT: RAHAOUI, HAKIM  
 ; TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES  
 ; FILE REFERENCE: BO43667-CIP  
 ; CURRENT APPLICATION NUMBER: US/09/995,587A  
 ; CURRENT FILING DATE: 2002-02-28  
 ; PRIOR APPLICATION NUMBER: 09/604,958  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: EPO 00201872.9  
 ; PRIOR FILING DATE: 2000-05-25  
 ; NUMBER OF SEQ ID NOS: 40  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 2394  
 ; TYPE: DNA  
 ; ORGANISM: Lactobacillus reuteri  
 US-09-995-587A-3

Query Match 3.8%; Score 52; DB 10; Length 2394;  
 Best Local Similarity 48.3%; Pred. No. 0.0032;

Matches 589; Conservative 0; Mismatches 535; Indels 96; Gaps 12;

QY 80 AAAACAGTCCCAATTTAAAGTCCCTCAATTCATCAGATCAGCAATCAAAAACATTGATT 139  
 DB 701 AACAGATGTCGGTACACAGTTCATCTTAAAGCAAGTAATCAAAAATATGCGCG 760  
 QY 140 CGGCAAAAGGTATATAGTACAGCACTTAATGA---TTAATGATGATGGATAGCT 196  
 DB 761 CGGCACTACCTAAAGTACCAAACTAATCTATGGAACCTTAAATGATGATGGATTAAT 820  
 QY 197 GGCACATCAAAAGCT---GATGCTAGTGGCGCAAAATATGATGATATACATGCT 253  
 DB 821 GGCAGTTCAGATGTTGGACAGCAAGTTCATATGATGATGATGATGATGATGATGATGAT 880  
 QY 254 CCGCTTTAGCAGGTGACCCAAAACAGTATGATCTCCACTCTATTTATCTATCAAA 313  
 DB 881 TCGCATATGATGGAATTCACAAACAA---ATGATATATCATCTATCTCTATATATA 937  
 QY 314 AAGTGTGATATCATCATGATGACAGTGAAGATGATGATGATGATGATGATGATGATGAT 373  
 DB 938 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 997  
 QY 374 ATAAATTTGTTCCAAATGATCCGTATCTTAATATCAAAACAGAGAGTGGTCAAGTCTG 433  
 DB 998 CTACCGCGGT-----TCACAAAGTGGTCAAGTCTG 1030  
 QY 434 CTACTTTAACCAAGATGGCAAGTCCGTTATCTATACAGATGATGATGATGATGATGATGAT 493  
 DB 1031 CTGTTTGAAGAGTATCACTATCTATCAATATTTATATA-----AGGTAG 1078  
 QY 494 AAGATGTGGAACCGGTGCTGTATCAAAATCATTTCAAGTCCCTCAAGTAACTATCCC 553  
 DB 1079 ACACGCTCTGATATACATACCAATCATCAAAAATGTCAGCGCTCTCTTTATTAACGTG 1138  
 QY 554 AGCCGATGACAGTCACTTAAGTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 613  
 DB 1139 ATAAATGAATGATATACCTCGCTCAGTACGAAATGACATAT-----TTGATTTTG 1192  
 QY 614 GCGGAGAGGTACATTTATCAAAATATTCAGCAATTTATGATGATGATGATGATGATGATGAT 673  
 DB 1193 AAGGTGATGCTATTTACTACCAAACTTATGATCAATGGAAGCTACTCAAAAG---TG 1249  
 QY 674 CAGGTGATACATCTTTAAGAGCCCTCACTATGATGATGATGATGATGATGATGATGATGAT 733  
 DB 1250 CCGATATATTTGCAATGCGGTGATGCTCATGATATGGAAGTGTATGATGATGATGATGATGAT 1309

QY 734 TTGTCTTGAACGATATCTGAAACAGATGTTATCAAGCGCATGCTTTCAATA 793  
 DB 1310 TTGTTTTCACCAAGTACTG---TTTGAAATATTCACAGGACCAACCAATTTATA 1366  
 QY 794 ATAAAGCTTACTATGCGGAGTACGCTCTCTCCAGATGAAAAAATAAAGTCTTC 853  
 DB 1367 ACTGGTTAAATATATGCGGAGATGACGCTTAATATCAAGAGCTTATTAATTAATCTTT 1426  
 QY 854 AAGTCTTAAAAACAAAT---GCTCTTGAAGATGATGATGATGATGATGATGATGATGAT 910  
 DB 1427 CCAATGATGATATTAAGAGTGGGCACTTGGCTAAATGACATATCGGTATCCCAAC 1486  
 QY 911 TGGCCGATGACTA-----TACAGCAAAAGTGTATGAACCATTAAGCGCATCA 961  
 DB 1487 TAAATAGACCAAAAGATCTTAAGTGGGAGATATATACACATTAATTTCTGCAC 1546  
 QY 962 ACACAGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1021  
 DB 1547 CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1606  
 QY 1022 TATTCAGGATTCAGAGATTCACAAATGATGATGATGATGATGATGATGATGATGATGAT 1081  
 DB 1607 TATTTGCGGCTACCGTTTAAATCGAGGAAGTAAATGATGATGATGATGATGATGATGAT 1666  
 QY 1082 AT-----ATGCTAGGCGCGGAGCGCATCTTAATATGCGCCAC 1120  
 DB 1667 ATGCGGTGATATATGTTGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1726  
 QY 1121 ACAACCGATTAATGAACACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1180  
 DB 1727 ATAGCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1786  
 QY 1181 CACACACTTACTCTATGCGGTATCCGACCTG---AAGTAAATATGATGATGATGATGAT 1237  
 DB 1787 CAGCAACTTATCATATTAATGCTGTCGCCGTTGCGGAAAAAGATGACCAATATTAAT 1846  
 QY 1238 CAAGTATATGACGAATAGA 1257  
 DB 1847 CTTCATATATGACTAATAGA 1866

RESULT 15  
 US-09-995-587A-4  
 ; Sequence 4, Application us/09995587A  
 ; Patent No. US2002127681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN HIJUM, SACHA ADRIANUS FOKKE TACO  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDIJNA HENDRIKA  
 ; APPLICANT: DIJKHUIZEN, LUBBERT  
 ; APPLICANT: RAHAOUI, HAKIM  
 ; TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES  
 ; FILE REFERENCE: BO43667-CIP  
 ; CURRENT APPLICATION NUMBER: US/09/995,587A  
 ; CURRENT FILING DATE: 2002-02-28  
 ; PRIOR APPLICATION NUMBER: 09/604,958  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: EPO 00201872.9  
 ; PRIOR FILING DATE: 2000-05-25  
 ; NUMBER OF SEQ ID NOS: 40  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 2592  
 ; TYPE: DNA  
 ; ORGANISM: Lactobacillus reuteri  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(51)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (68)..(2434)  
 US-09-995-587A-4

Query Match 3.8%; Score 52; DB 10; Length 2592;  
Best Local Similarity 48.3%; Pred. No. 0.0034;  
Matches 589; Conservative 0; Mismatches 535; Indels 96; Gaps 12;

QY 80 AAAACAGTCTCAATTTAAAGTCCATTTCAATGCAATCAGCAATCAAAAACATTGATT 139  
DB 741 AACAGATGCTGGTACACAGATTCATTTAAAGCAGTGAATCAAAAATATGCGCG 800  
QY 140 CGGCAAAAAGGTATGATGAGTACAGCACTTAATAGA---TTAGATGATGGATAGCT 196  
DB 801 CCGCTACACACTAAAGATGCAAACTAATACTATTGAACTTTAGATGATGGATTCAT 860  
QY 197 GGCACCTGCAAAACGCT---GATGGTACTGCGCAATTTATCATGATATCATGCTCT 253  
DB 861 GGCACCTGCAAAAGTCTGCGACAGCAAGTGTATGATGATGCTATCAACTGTCA 920  
QY 254 CCGCTTTACAGGTGACCCAAAACAGATGATGATCTCACTTCTTTATCTATCAAA 313  
DB 921 TCGCAATGATGGAAATTCCAAAACAAA---ATGATTAATATCTATCTATTAATA 977  
QY 314 AAGTCGGTATCATGATGATGACAGTGGAAAAATGCTGAAGATATTGAAGATATGG 373  
DB 978 AGTATGATGATATGATGATTAATGATCAATGAGATGATGATGATGATGATGAT 1037  
QY 374 ATAAATTGTTCCAAATGATCCGATCTTAATATCAAAACAGAGATGCTCAGTTCTG 433  
DB 1038 CTACCGCGGCT-----TCACAGAAATGCTCAGATCAG 1070  
QY 434 CTACCTTTACCAAGATGCGCAAGTCCGTTTATGATGATGATGATGATGATGAT 493  
DB 1071 CTGTTTTCAGAGTGTAACTTATCCAAATTTATTAACA-----AGGGTAG 1118  
QY 494 AAGATGATGGAACCGGTGCTGTTAACCAAAATCATTTCAACTGCTCAAGTAACTTATCCC 553  
DB 1119 ACACGCTGTATGATACATACCAATCAATCAAAATTTGCTAGGCTACTCTTTATTAACG 1178  
QY 554 ACCCGGATGCACTACCTTAAAGTGTGATGATGATGATGATGATGATGATGATG 613  
DB 1179 ATAAATATGGAATGATCACTGCTCAGGTACGAATGACTATA-----TTGATTTG 1232  
QY 614 GCGGAGAGGTTACAGTTTATCAAAATATTCAGCAATTTATGATGATGATGATGAT 673  
DB 1233 AAGGTGATGCTATTACTACCAACTTATGATCAATGGAAGCTTACTAACAAG---TG 1289  
QY 674 CAGGTGATTAACCTACTTTAAGAGACCCCTCACTATGTTGAAGATTAAGGCGCATTAATATC 733  
DB 1290 CCGATAATATGCAATGCGTATGCTCATGATTAATGAATGATGATGATGATGATGAT 1349  
QY 734 TTGCTTTGAAGCAATACTGGAACAACAGATGTTATCAAGGCGATCAGTCTTCAATA 793  
DB 1350 TTGTTTTCACCAAGTACTGG---TTTGAAAAATTTATCAAGGCGAGCACCAAAATTTATA 1406  
QY 794 ATAAAGCTTACTATGCGGAAGTGAAGTCTTCTCCAGATGAATAAAATTAACGCTGC 853  
DB 1407 ACTGTTAATATTATGCGGAGATGAGCGCATTAATATCAAGAGCTTATTTAGAAATCTTT 1466  
QY 854 AAAGTCTTAAAAACAATTT---GCTTCTTTAGCGAATGCTCATTTAGCATTTGTAAT 910  
DB 1467 CCAATGATGATATTTAAGAGTGGGCACTTGGGCTAATGCACTATCGGTATCCTCAAC 1526  
QY 911 TGCGCGATGACTA-----TACAGTAAAAAGTGTATGAAACCATTTAGTCGATCA 961  
DB 1527 TAAATAGAGCAAAAAGATCCCTAAGTGGCAGAGTATATCTACCATTAATTTCTGCAC 1586  
QY 964 ACACAGTACAGATGATGATGAGCGCAATATATTTAATGATTAATTAATGATGATG 1021  
DB 1587 CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1646  
QY 1022 TATTCAGGATTCAGAGAGATCAAAATGACAGATGATGATGATGATGATGATGATG 1081  
DB 1647 TATTTGCGGCTACCGTTTAATTCAGAGAGATGATGATGATGATGATGATGATGAT 1706  
QY 1082 AT-----ATGCTAAGGCCCGAGCGAGCTCTTAAATGGCCAC 1120

DB 1707 ATGCCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1766  
QY 1121 ACAACCCGATTAATGAACTGAGCTTGTATTAACATGATGATGATGATGATGATG 1180  
DB 1767 ATAAAGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1826  
QY 1181 CACACACTTACTGCTATGCGGTATGCGGCAACCTG---AAGGTAAATGATGATGAT 1237  
DB 1827 CAGCACTTATTCATATTTATGCTGTCCCGTTGCGGGAAGATGACCAAGATTAAT 1886  
QY 1238 CAGTTAATGACGAATAGA 1257  
DB 1887 CTTCATATATGACTAATAGA 1906

Search completed: June 18, 2003, 15:23:44  
Job time : 211.939 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 23:20:16 ; Search time 301.364 Seconds  
(without alignments)  
10200.228 Million cell updates/sec

Title: US-09-986-682B-4

Perfect score: 1365

Sequence: 1 ATGACACGCGGCGCTACAA.....CCGACACAGCAATTCCTCCA 1365

Scoring table:

IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1365	100.0	2408	19 AAV17621
2	617.2	45.2	1449	24 ABK75353
3	577	42.3	10317	21 AA253988
4	575.4	42.2	1668	21 AA257329
5	575.4	42.2	1722	21 AA257330
6	542.8	39.8	1584	23 AAS68494
7	542.8	39.8	1935	23 AAS70417
8	542.8	39.8	2044	23 AAS5834
9	542.8	39.8	2044	23 AAS87384

10	542.8	39.8	2325	23 AAS69777	DNA encoding novel
11	542.8	39.8	2325	23 AAS75941	DNA encoding novel
12	542.8	39.8	2351	12 AAO10498	B. amy101liquefacien
13	542.8	39.8	2370	23 AAS68133	DNA encoding novel
14	542.8	39.8	2370	23 AAS92594	DNA encoding novel
15	542.8	39.8	2374	23 AAS87392	DNA encoding novel
16	542.8	39.8	2472	23 AAS89566	DNA encoding novel
17	542.8	39.8	2479	23 AAS86776	DNA encoding novel
18	542.8	39.8	2562	23 AAS88281	DNA encoding novel
19	542.8	39.8	2568	23 AAS6687	DNA encoding novel
20	542.8	39.8	2682	23 AAS71354	DNA encoding novel
21	542.8	39.8	2682	23 AAS87361	DNA encoding novel
22	542.8	39.8	2742	23 AAS66648	DNA encoding novel
23	542.8	39.8	2790	23 AAS87386	DNA encoding novel
24	542.8	39.8	3084	23 AAS85832	DNA encoding novel
25	542.8	39.8	3084	23 AAS87383	DNA encoding novel
26	542.8	39.8	3267	23 AAS87793	DNA encoding novel
27	542.8	39.8	3505	23 AAS85838	DNA encoding novel
28	542.8	39.8	4132	23 AAS86045	DNA encoding novel
29	542.8	39.8	5912	23 AAS86048	DNA encoding novel
30	542.8	39.8	5938	23 AAS87395	DNA encoding novel
31	542.8	39.8	6721	24 AAS18599	PurIntergic recepto
32	542.8	39.8	6721	24 AAS18600	PurIntergic recepto
33	542.8	39.6	1690	23 AAS85792	DNA encoding novel
34	541.2	39.6	1690	23 AAS86019	DNA encoding novel
35	541.2	39.6	1690	23 AAS87350	DNA encoding novel
36	541.2	39.6	3288	23 AAS73395	DNA encoding novel
37	541.2	39.6	3288	23 AAS85789	DNA encoding novel
38	541.2	39.6	3288	23 AAS86010	DNA encoding novel
39	541.2	39.6	3288	23 AAS87346	DNA encoding novel
40	541.2	39.6	3288	23 AAS88746	DNA encoding novel
41	539.6	39.5	2328	23 AAS64228	DNA encoding novel
42	539.6	39.5	3305	18 AAT68805	Mutant levanucras
43	532.8	39.0	1629	23 AAS85800	DNA encoding novel
44	532.8	39.0	1629	23 AAS86022	DNA encoding novel
45	532.8	39.0	1629	23 AAS87359	DNA encoding novel

# ALIGNMENTS

RESULT 1	
AAV17621	AAV17621 standard; DNA; 2408 BP.
ID	
AC	AAV17621;
XX	
DT	04-JUN-1998 (first entry)
DE	Bacillus sp. V230 beta-fructofuranosidase genomic DNA.
XX	
KW	Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
KW	lactosucrose; fructosyltrehalose; anticariogenic; growth promoter;
KW	Bifidobacteria; mineral/absorption promoting activity; taste; texture;
KW	cosmetic; pharmaceutical; sweetener; ds.
OS	Bacillus sp.
XX	
FH	Key
FT	5'UTR
FT	Location/Qualifiers
FT	1..360
FT	/*tag= a
FT	361..456
FT	/*tag= b
FT	457..1821
FT	mat_peptide
FT	/*tag= c
FT	/EC-number= 3.2.1.26
FT	/product= beta-fructofuranosidase
FT	/note= "Claim 7"
FT	1822..2408
FT	/*tag= d
XX	
PN	EP812915-A2.

PD 17-DEC-1997.  
XX 10-JUN-1997; 97EP-0304032.  
XX 10-JUN-1996; 96JP-0170630.  
XX (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
PI Chaeen H, Kubota M, Tsusaki K;  
XX WPI: 1998-034976/04.  
DR P-PSDB: AAM48306.  
XX Bacillus beta-fructofuranosidase enzyme - useful for producing  
PT fructo:furanosylated saccharide(s) or alcohol(s)  
XX Claim 8; Pages 23-25; 32pp; English.  
PS This is a genomic DNA sequence which encodes  
CC beta-fructofuranosidase, an enzyme that has an optimum temperature  
CC of 50 deg. C at pH 6. The enzyme was produced by recombinant DNA  
CC technology. It catalyses fructofuranosyl transfer reactions from a  
CC fructofuranosyl donor to a fructofuranosyl acceptor where the donor is  
CC sucrose, raffinose or erlose and the acceptor is selected from alcohols,  
CC sugar alcohols and saccharides having no beta-fructofuranosidic linkages,  
CC especially where the reaction product is xylosyl fructoside, erlose,  
CC isomalosyl fructoside, lactosucrose or fructosyltrehalose.  
CC Such reaction products have a satisfactory taste and sweetness, a  
CC moderate viscosity and humectancy, an effective anticariogenic activity,  
CC growth promoting activity for Bifidobacteria, mineral-absorption  
CC promoting activity. The products can be used to improve the tastes and  
CC textures of food products, cosmetics and pharmaceuticals and are useful  
CC as sweeteners in the food and pharmaceutical industries.  
XX  
SQ Sequence 2408 BP; 777 A; 492 C; 496 G; 643 T; 0 other:  
  
Query Match 100.0%; Score 1365; DB 19; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
1 ATGAACACGGGGGACTACAGAGAGACTATGTTTGGCCATATACACGGCGTACATG 60  
DB 457 ATGAACACGGGGGACTACAGAGAGACTATGTTTGGCCATATACACGGCGTACATG 516  
QY 61 CTAAAAATTCAGAGCAACAACAGTCTCAATTTAAAGTGCCTCAATTCATGATCA 120  
DB 517 CTAAAAATTCAGAGCAACAACAGTCTCAATTTAAAGTGCCTCAATTCATGATCA 576  
QY 121 GCATCAAAAACATTGATTCGGCAAAAGGATATGTAAGTCAGCGCACTTAATAGATTTA 180  
DB 577 GCATCAAAAACATTGATTCGGCAAAAGGATATGTAAGTCAGCGCACTTAATAGATTTA 636  
QY 181 GATGATGAGGATGATGCGCACTGCAAAAGCTGATGTACTGCGGCAAAATATCATGGA 240  
DB 637 GATGATGAGGATGATGCGCACTGCAAAAGCTGATGTACTGCGGCAAAATATCATGGA 696  
QY 241 TATCAACATCGTCTCGCTTTAGCAGGTGACCAAAAACAGTGAATGATCTCCACTTAT 300  
DB 697 TATCAACATCGTCTCGCTTTAGCAGGTGACCAAAAACAGTGAATGATCTCCACTTAT 756  
QY 301 TTATTCATCAAAAAGTGGTGATATCATGATTCAGACCTGGAAAAATGCTGGAAGATTA 360  
DB 757 TTATTCATCAAAAAGTGGTGATATCATGATTCAGACCTGGAAAAATGCTGGAAGATTA 816  
QY 361 TTGAGAGTATGATGATTTGTCCAAATGATCCGTTCTTAATATCAACACAGAGAG 420  
DB 817 TTGAGAGTATGATGATTTGTCCAAATGATCCGTTCTTAATATCAACACAGAGAG 876  
QY 421 TGTGTCAGGTTCTGCTACTTTAAACCAAGATGCCCAAGCTTTATTTATATACAGATTAC 480  
DB 877 TGTGTCAGGTTCTGCTACTTTAAACCAAGATGCCCAAGCTTTATTTATATACAGATTAC 936  
QY 481 TCAAGTAACTCTGAAGATGTGGAACCGGTCTGTGTAACCAAAATCATTTCAATGCTCTCA 540

DB 937 TCAGGTATCTCTGAAGATGTGGAACCGGTCTGTGTAACCAAAATCATTTCAATGCTCTCA 996  
QY 541 GTAACTTATCCCGACCGGATGACGCTACCTTAAAGTCGATGATCTGATCATATAA 600  
DB 997 GTAACTTATCCCGACCGGATGACGCTACCTTAAAGTCGATGATCTGATCATATAA 1056  
QY 601 TCTGCTTTGATGGCGGAGACGCTACAGTTATCAAAATATTCGCAATTTATGATGAA 660  
DB 1057 TCTGCTTTGATGGCGGAGACGCTACAGTTATCAAAATATTCGCAATTTATGATGAA 1116  
QY 661 GGCAAGTGGATTTAGGTGATTAACCATTTTAAAGACCTCTACTATGTTGAAGATPAAG 720  
DB 1117 GGCAAGTGGATTTAGGTGATTAACCATTTTAAAGACCTCTACTATGTTGAAGATPAAG 1176  
QY 721 GGCCATTAATATCTTGTCTTTGAAGCGAATCTGGAACACAGATGTTTACAGCGCAT 780  
DB 1177 GGCCATTAATATCTTGTCTTTGAAGCGAATCTGGAACACAGATGTTTACAGCGCAT 1236  
QY 781 CAGTCTTCATTAATTAAGCTTACTATGGCGGAGTGCCTTCTCCAGAAATGAATAA 840  
DB 1237 CAGTCTTCATTAATTAAGCTTACTATGGCGGAGTGCCTTCTCCAGAAATGAATAA 1296  
QY 841 AATAAAGTCTTCAAAAGTCTTAAACCAAAATTCCTTTAGCGAATGTTGATTAAGGC 900  
DB 1297 AATAAAGTCTTCAAAAGTCTTAAACCAAAATTCCTTTAGCGAATGTTGATTAAGGC 1356  
QY 901 ATTGTGATTAATGGCGGATGATATTAACGTAAGAGTGTATGAACCAATTAAGTGCATCA 960  
DB 1357 ATTGTGATTAATGGCGGATGATATTAACGTAAGAGTGTATGAACCAATTAAGTGCATCA 1416  
QY 961 AACACAGTACAGATGAAGTGAACGCGCAATATTTAAATGAATTAATTAATGATAT 1020  
DB 1417 AACACAGTACAGATGAAGTGAACGCGCAATATTTAAATGAATTAATTAATGATAT 1476  
QY 1021 CTATTCACGATTCACAGATTCACAAATTAATGAATGAATTAATTAATGATAT 1080  
DB 1477 CTATTCACGATTCACAGATTCACAAATTAATGAATGAATTAATTAATGATAT 1536  
QY 1081 TATATGCTAGGCGCGGAGGAGTCTTAAATGAGCCACACCAACCGGATTAATGAAT 1140  
DB 1537 TATATGCTAGGCGCGGAGGAGTCTTAAATGAGCCACACCAACCGGATTAATGAAT 1596  
QY 1141 GGACTTGTATGAACATGAATCTTGACCTCTGATGTCACACACACTTACTCTCATTTGC 1200  
DB 1597 GGACTTGTATGAACATGAATCTTGACCTCTGATGTCACACACACTTACTCTCATTTGC 1656  
QY 1201 GGTATCCCGCACCTGGAAGTAAATATGTTGTTACTCACAGTTATATGACGAATAGAGGC 1260  
DB 1657 GGTATCCCGCACCTGGAAGTAAATATGTTGTTACTCACAGTTATATGACGAATAGAGGC 1716  
QY 1261 TTCTATCCAGAAACATCTGACCTGCGGAGCAAAAGCTTGGGGTTAATTAATGAAGGGCT 1320  
DB 1717 TTCTATCCAGAAACATCTGACCTGCGGAGCAAAAGCTTGGGGTTAATTAATGAAGGGCT 1776  
QY 1321 GACACATCTGGAGGAGAAATAGTTCCGAGCAAGAGCAATTTCCCA 1365  
DB 1777 GACACATCTGGAGGAGAAATAGTTCCGAGCAAGAGCAATTTCCCA 1821  
  
RESULT 2  
ABK75353  
ID ABK75353 standard; DNA: 1449 BP.  
XX  
XX ABK75353;  
XX  
XX 13-AUG-2002 (first entry)  
XX  
XX DE Bacillus licheniformis genomic sequence tag (GST) #2644.  
XX  
XX KM Differential gene expression; genomic sequenced tag; GST;  
KM altered culture condition; environmental stress;  
KM physiological provocation; ds.

XX Bacillus licheniformis.  
 XX OS  
 XX MO200229113-A2.  
 XX PN  
 XX 11-APR-2002.  
 XX PD  
 XX 05-OCT-2001; 2001WO-US31437.  
 XX PF  
 XX 06-OCT-2000; 2000US-0680598.  
 XX PR 27-MAR-2001; 2001US-279526P.  
 XX PA (NOVO ) NOVOZYMES BIOTECH INC.  
 XX PA (NOVO ) NOVOZYMES AS.  
 XX PI Berka R, Clausen IG;  
 XX WPI: 2002-416684/44.  
 XX DR  
 XX Monitoring differential expression of several genes in first Bacillus  
 PT cell relative to expression of same genes in one or more second  
 PT Bacillus cells, by using substrate containing Bacillus genomic  
 PT sequenced tag array -  
 PS  
 PS Claim 4; SEQ ID NO 2644; 200pp; English.  
 XX  
 CC The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridizing labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions.  
 CC environmental stress or other physiological provocation. Extensive  
 CC follow-up characterisation is unnecessary, when one spot on an array  
 CC equals one gene or one open reading frame, since sequence information is  
 CC available. This sequence represents a genomic sequence tag (GST) used in  
 CC the method of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at  
 CC ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 CC Sequence 1449 BP; 492 A; 341 C; 313 G; 303 T; 0 other;  
 SQ  
 Query Match 45.28; Score 617.2; DB 24; Length 1449;  
 Best Local Similarity 67.38; Pred. No. 3.7e-162;  
 Matches 896; Conservative 0; Mismatches 418; Indels 18; Gaps 1;

Db 339 TTTGCGCTGGCGGGCATCCGAAAGACGTGATGACATCCATCTATTGTTCTATCA 398  
 Qy 312 AAAAGTGGTATATCATTCGATTGACAGCTGGAAGAAATGCTGGAGAGATATTGAGATAT 371  
 Db 399 AAAAGAAAGCGAAACTTCTATGACAGCTGGAAAGCCGGCGAGTGTTTAAAGACAG 458  
 Qy 372 GGATAAATTTGTCCAAATGATCCGTACTTAAATATCAACACAGAGAGTGTGAGCTTC 431  
 Db 459 CGACAAATTTGTTCACAGCATCCGTACTTCAACATCAACACAGAGATGTAGCTTC 518  
 Qy 432 TGCTACTTTTAAACAAAGATGGCCAGTCCGTTTATCTATACAGATTACTAGGTAATCC 491  
 Db 519 TGCACCGCTGACAAAGACGAAAGATGCCAGCTGTTTAAACAGCTTTTCCGCGACGCA 578  
 Qy 492 TGAAGATGTGGAACCGGTGCTGTTAAACCAATATATTCAACTGCTCAAGTAACCTATC 551  
 Db 579 ATA-----CGGCAAGCACAGCCTGACAAACAGCTCAGTCAATTTCTC 620  
 Qy 552 CCAGCCGGATGCAGCTACACTTAAAGTCGATGAGATCTGATCATTAATCTGCTTGA 611  
 Db 621 TCAGCCGGATTCGGACACGCTCAAAATTTAGCGGTGAGAGATCATTAATGCTCTTTGA 680  
 Qy 612 TGGCGAGACGCTACAGTTTATCAAAATATTACAGCAATTTATCGATGAGGCAAGTGGAT 671  
 Db 681 CGGCGCGAGCGCAACGGTATACCAAAAGCTTCACCAATTCATGACGAAGCAACTACAG 740  
 Qy 672 TTCAAGGTATACCATATCTTTAAAGACCCCTCACTATGTTAAGTAAGGCCCTAATA 731  
 Db 741 CTCGGCGCACCAACCATATGATGAGAGACCCCATTAATGTGAAGACCGCCCATTAATA 800  
 Qy 732 TCTGTCTTTAAGCAATATCTGAAACAAAGATGTTATGAGCGCATGCTTTTCAA 791  
 Db 801 TCTGTATTTTAAGCAATATCTGAAACAAAGATGTTATGAGCGCATGCTTTTCAA 860  
 Qy 792 TAATAACTTACTATGCGGAGTGAAGTCTTCTTCAGAAATGAAGAAATTAATAGCT 851  
 Db 861 CAACAGACCTACTAGCGGGGACAGAAAGTCTTTAAAGAAAGCAAGCAGACGCTGCT 920  
 Qy 852 TCAAGTCTTAAAGAAACAAATGCTCTTTAGCCAAAGTGTGATGAGCATTTGTTGAT 911  
 Db 921 GCAAGGTGCGCAAAAGAAAGAGCTGCTGCGGACCGCCCTCGGAATCATGCAAT 980  
 Qy 912 GCGCGATGATCATACAGGAAAGTGTATGAACCATTAAGTCATCAACACAGTATG 971  
 Db 981 AATATACGATTAATACACTGAAAGAAAGTATGAAGCTTTGATCGCTTCAATFACGTTGAC 1040  
 Qy 972 AGATGAAGTGCAGACGCGCAATATATTTAAATGAATTAATGATCTATTCACAGGA 1031  
 Db 1041 AGATGAATGCAGACGCGCAATATATTTAAATGAATGAATGATCTATTCACAGGA 1100  
 Qy 1032 TTCAAGAGATCCAAATGACGAGTGAATGATTAACGACAAAGATGTTATATGCTAGG 1091  
 Db 1101 TTCAAGAGATCCAAATGACGAGTGAATGATTAACGAGCTTTTAAAGCAATTTATGCTAGG 1160  
 Qy 1092 GCCCGAGAGCAGCTTAAATGAGCCACACAGCCGATTAATGAATGAATGATCTGAT 1151  
 Db 1161 CTATGATACAGCTTATTAACCGGACCAATTAAGCTTTTAAAGCAATTCGAGCTGTTT 1220  
 Qy 1152 GAACATGAATCTTGAACCTGATCTACACAGCACTACTACTGATGCGGTATCCGCA 1211  
 Db 1221 GCATATGAGACGAGATTAACATGACATCACTGTTACTTATTCACACTTTGCCGATCCGCA 1280  
 Qy 1212 CCGTGAAGGTATATATGCTGATCACAAGTATATATGACAAATGAGAGCTTTATCCAGA 1271  
 Db 1281 GAAAGAAAGGCGACAGATGCTATACAAAGTATACAAAGCAAGAGGATTTGAAGCA 1340  
 Qy 1272 ACATCACTCTACCTGCGGAGCAAGCTTGGGTTAATATTAAGAGGCTGTGACATCGG 1331  
 Db 1341 GCATCAGCCACGTTTGCACCAAGCTTTTGTCTGAAGATCAAGAGATCAAAACATCCTGT 1400  
 Qy 1332 AGGAGAAATATAG 1343

Db 1401 TGTCAAAACAG 1412

RESULT 3  
AAZ35988/c  
ID AAZ35988 standard; DNA: 10317 BP.  
XX  
AC AAZ35988;  
XX  
DT 09-FEB-2000 (first entry)  
XX  
DE Vector pGendell sequence SEQ ID NO:1.  
XX  
XX DNA sequencing vector; nested deletion; transposition; contig;  
KM truncated lacZ; origin of replication; pGendell; mapping; ss.  
XX  
OS Synthetic.  
XX  
PN MO9953044-A2.  
XX  
PD 21-OCT-1999.  
XX  
PF 09-APR-1999; 99WO-1B00690.  
XX  
PR 10-APR-1998; 98US-0058746.  
XX  
PA (GEST ) GENSET.  
XX  
PI Chumakov I, Tanaka H;  
XX  
DR WPI: 2000-023040/02.  
XX  
PT New DNA sequencing vectors, used for sequencing large regions of DNA  
PT and mapping the location of markers -  
XX  
PS Claim 43; Page 93-96; 102pp; English.  
XX  
XX The present invention describes a vector comprising the following:  
CC (a) a high copy number origin of replication (Ori) having at least one  
CC cloning site, with at least one cloning site being positioned in the  
CC high copy number Ori such that the ability of the high copy number Ori  
CC to direct replication is not disrupted when no insert has been cloned  
CC into the cloning site; and (b) a low copy number Ori. Also described is a  
CC truncated lacZ gene to distinguish cells with the truncated lacZ  
CC gene at high copy number from cells carrying it at a low copy number,  
CC where the former are dark blue when grown on medium containing Xgal and  
CC IPTG and the latter are light blue when grown on medium containing Xgal  
CC and IPTG. Products and methods from the present invention can be used  
CC for the manipulation of DNA. They can be used for sequencing large  
CC regions of DNA and mapping the locations of markers within large regions  
CC of DNA. The present sequence represent a specifically claimed vector  
CC  
XX  
XX Sequence 10317 BP: 2567 A; 2296 C; 2469 G; 2985 T; 0 other;  
S0

Query Match 42.3%; Score 577; DB 21; Length 10317;  
Best Local Similarity 67.2%; Pred. No. 1.5e-150;  
Matches 902; Conservative 0; Mismatches 390; Indels 51; Gaps 4;

QY 1 ATGAACAGCGGGAGCTACAGAGAGACTAGTGTGGCCATATTATACAGCGGCTGACATG 60  
DB 8707 ACGAAGCAAAAGCCATATTAAGAAACATACGCAATTCCTCCATATTACAGCGCATGATATG 8648  
QY 61 CTTAAATTCAGAGACACAAACAGTCTCAATTTAAAGTGCCCTCAATTAATGATCA 120  
DB 8647 CTGCAAAATCCTGGAACGCAAAAATGAAAAATATCAAGTTCCTGAGTTGATTCGCTCC 8588  
QY 121 GCAATCAAAAACATGATGGGCAAAAGGATGATGATAGTCAGCAACTTAATGATTTA 180  
DB 8587 ACAATTAATAATATCTCTTCTGCAAAAG-----CCTG 8555  
QY 181 GATGTATGGATACCTGGCCACTGCAAAACCGCTGATGTACTGGCAAAATTAATCATGA 240

Db 8554 GACGTTTGGAGCACTGGCCATTACAAAAGCTGACGGCACTGTGCAAACTATACAGGC 8495  
QY 241 TATCATATCGTCTCCGCTTTAGCAGGTGACCCAAAACAGTATGATCTCCACTTCAT 300  
DB 8494 TACCACATTCGCTTTGGATTAGCCGGAGATCTTAAAAATGGGATGACATCATCTTAC 8435  
QY 301 TTATTCATCAAAAAGTCGGGTGATACATGATGACAGCTGGAAAAATGGCTGGAAGATA 360  
DB 8434 ATGTTCTATCAAAAAGTCGGGAAACTTCTATGACAGCTGGAAAAACCTGGCCGCTC 8375  
QY 361 TTTGAGATATGATTAATTTGTTCCAAATGATCCGATCTTAATATCAACACAGAG 420  
DB 8374 TTTAAAGACAGCACAATTCGATGATGATGATTCATTCCTAAAGACCAACACAGAA 8315  
QY 421 TGTGACGTTCTGCTACTTTTACCAGATGGCCAGTCGTTTATCTATACGATTAC 480  
DB 8314 TGTGACGTTCTGACACATTTACATCTGACGGAAAAATCCGTTTATCTACACTGATTTCC 8255  
QY 481 TCAGTATATCCCTGAAGATGGTGGAAACCGGTGCTGTAAACCAATTCATTCAGCTCAA 540  
DB 8254 TCCGTTAAACATTA-----CGGCAAAACAACTGACACTGACAA 8213  
QY 541 GTAACCTATCCAGCCGATGACACTTAAGTCGATGAGATGATCATATAA 600  
DB 8212 GTTAACGATACAGCATCAGACAGCTCT--TTGAACATCAACGGGTGATAGAGATTATAA 8156  
QY 601 TCTGCTTTGATGGCGGAGACGTTACAGTTATCAAAATATTCAGCAATTTATGATGAA 660  
DB 8155 TCAATCTTTGA--CGGTGACGGAAAAACGATCAAAATGATGACAGCTATCATGATGAA 8099  
QY 661 GGCAAGTGATTTGAGTGATTAACCATTAAGTGAAGCCCTCACTATGTTGAAGATAAG 720  
DB 8098 GGCAACTACAGCTGACGGCAGCAACATACGCTGAGAGATCTCTACGTAGAAAGATTA 8039  
QY 721 GGCCATAATATATCTTCTTTGAAGCAATACGTAACCAACAGATGTTATCAAGCGAT 780  
DB 8038 GGCCACAATATCTAGTATTGGAAGCAAACTGCACTGAAAGTGGCTACCAAGCGGAA 7979  
QY 781 CAGCTTTCAATTAATTAAGCTTACTATGCGGAGTACGCTTCTTCCAGATGAATAA 840  
DB 7978 GAACTTTTATTAATTAACAAAGCTACTATGCGAAAGCAACATCTTCTCCGCAAGAAGT 7919  
QY 841 AATTAACGCTTCTCAAGCTCAAAACAAATTCCTTTAGAGATGGTCAATTAAGC 900  
DB 7918 CAATAACTCTGCAAAAGGATTAATAAAGCAAGCGCTAGTTAGAAACGGGCTCTCGGT 7859  
QY 901 ATGTGTAATGGCCGATGACTATACAGTGAAGAGTATGAACAATTAATGTCATCA 960  
DB 7858 ATGATTGAGTAAACGATGATTAACACACTGAAAAAGTGAATGAACCGCTGATTCATCT 7799  
QY 961 AACACAGTACAGATGAAGTGAAGCGGCCAATTAATTAATGAATTAATTAATGATAT 1020  
DB 7798 AACACAGTACAGATGAAGTGAAGCGGCCAATTAATTAATGAAGCAAAATGATAT 7739  
QY 1021 CTATTACAGGATTAAGAGATCCAAATGACAGATGATGAATTAACGACAAAGATGTT 1080  
DB 7738 CTGTTCACTGACTCCCGGATCAAAATGACATGATGACGCGATTAAGTCAACGATAT 7679  
QY 1081 TATATGCTAGGGCCCGGAGCGGACTCTTAATATGGCCACAAACCCGATATGAAT 1140  
DB 7678 TACATGCTGGTGTATGTTTCTTAATTTTAACTGGCCATCAAGCCGCTGACAAACT 7619  
QY 1141 GCACTTGATTAAGATGAATCTTGACCCCTGCTATCTCAACACACTTACTCTCATTCG 1200  
DB 7618 GGCCTTGTTTAAAAATGATGATGATTCATGACATGATTAACCTTACTCTACCTTC 7559  
QY 1201 GGTATCCGCGACCCCTGAAGATTAATATGTTGATCTACAAGTATATAGCAATAGAGGC 1260  
DB 7558 GCTGTACTCAAGGAAAGAAACATATGCTGATTTACAGCTATATAGCAACAGAGGA 7499  
QY 1261 TTCTATCCAGAACATCACTCTACCTGCGGACAGGCTGGGGTTAATTAATTAAGGGCTCT 1320

Db	7498	TTCTACGAGACAAACATCAACGTTTGGCGGACCTTCCTGCTGACATCAAGGCAAG	7435
Oy	1321	GACACATCTGGAGAGAAAATAG	1343
Db	7438	AAACATCTGTTGCAAGACAG	7416
RESULT 4			
AA257329	AA257329 standard; DNA; 1668 bp.		
AA257329	AA257329;		
03-APR-2000	(first entry)		
SacB and cpy chimeric gene #1.			
levansucrase; SacB; cpy; salt tolerance; vacuole guide peptide;			
carboxypeptidase A; chimeric gene; transgenic plant; yeast;			
Bacillus subtilis 168; Saccharomyces cerevisiae X8;			
drought resistance; ds.			
Chimeric - Saccharomyces cerevisiae.			
Chimeric - Bacillus subtilis.			
CN1231337-A.			
13-OCT-1999.			
08-APR-1998;	98CN-0101336.		
08-APR-1998;	98CN-0101336.		
(GENE-) INST GENETICS CHINESE ACAD SCI.			
Chen S, Zhang H, Dong W;			
WPI: 2000-087902/08.			
Method for improving plant salt resistance using transgenic technology			
Claim 6; Page 2; 12pp; Chinese.			
A method has been developed for raising salt tolerance of plants using transgenic technology. The method comprises: cloning levansucrase gene SacB and vacuole guide peptide (cpy) sequence of carboxypeptidase A from Bacillus and DNA of yeast, using the two kinds of gene to create a chimeric gene, and using the chimeric gene to structure plant expression carrier, using the obtained plant expression carrier to transform a plant and screen for resistant seedlings. The method can obtain drought-resistant salt-tolerance plants, and the polymerase chain reaction (PCR) and Northern analysis of transgenic plants shows that the exogenous gene is integrated in the salt-tolerance transgenic plant. The present sequence represents a SacB and cpy chimeric gene from the present invention.			
SQ	Sequence	1668 bp; 578 A; 372 C; 325 G; 393 T; 0 other;	
Query Match	42.2%;	Score 575.4;	DB 21; Length 1668;
Best Local Similarity	67.1%;	Pred. No. 1.9e-150;	
Matches	901; Conservative	0; Mismatches 391;	Indels 51; Gaps 4;
Oy	1	ATGAACACCGGGAGCTACAGAGACTATGCTTTGGCCATATTACAGCGGACATG	60
Db	340	ACGAAACAAAAGCCATATAAGAGAAACATRGCGCATTTCCATATTACAGCGCATGATG	399
Oy	61	CTAAAAATTCGAGGACAAACAACAGTCTCAATTAAAGTCCGCTCAATTCATGATCA	120
Db	400	CTGCAAAATCCCTGAACACGCAAAAATGAAAATATCAAGTTCCGATTCGATTCGTC	459
Oy	121	GCATTCAAAACATGATGTTGGCGAAAAGGATATGATGAAGTCAAGCACTTATAGATT	180
Db	460	ACAATTAATAATCTCTTCTGCAAAAG-----CCTG	492

QY	101	GATGATATGGGATAGCTGGCCACTGCAAAACGCTGATGGTACTGGCGCAATATATCATGGA	240
Db	493	GACGTTTGAGACAGCTGGCCATTACAAAACGCTGACGGCAGCTGTGCAAACTATCACGGC	552
QY	241	TATCACTAGCTGTCCGCTTACAGGTGACCCCAAAAACAGATGATATGCTCATTCAT	300
Db	553	TACCACTATCGTCTTTGTCATTAGCCGAGATCCTAAAAATGGCGATGACACATCGATTTAC	612
QY	301	TTATTCATCAAAAAGTGGGTGATATCATGATTTGACAGCTGGAAAAATGCTGAAGACTA	360
Db	613	ATGTTCTATCAAAAAGTGGGCAAACTCTATTTGACAGGTGAAAAACGCTGGCGCGTC	672
QY	361	TTTGAAGATATGATAAATTTGTCCAAATGATCCGTATCTTAAATATCAACACAGAG	420
Db	673	TTTAAAGACAGGACAAAATTCGATCCAAATGTATTTATCTTAAAGACCAAAACACAAAG	732
QY	421	TGCTCAGGTTCTGCTACTTTAAACAAGAATGGCAAGTCCGTTATTTGATACATTAAC	480
Db	733	TGCTAGGTTACGCCACATTTTACATCTGACGGAAAAATCCGTTTATTTCTACACTATTTTC	792
QY	481	TCAGSTAATCCTGAAGATGTGGAAACGGTGTGTTAACCAAAATCTTCACTGCTCAA	540
Db	793	TCCGTTAAACATTA-----CGCAAAACAACACTGCACCTGCACAA	834
QY	541	GTAACCTATCCCAACCGGATCCAGCTACACTTAAAGTCGATGGAGTATTCATATATA	600
Db	835	GTTACGATATCGCATCAGCATCAGACGCTCT--TTGAACATCAACGGGTGTAGAGATTATATA	891
QY	601	TCGTCTTTGATGGCGGAGACGGTACAGTTTATCAAAATATTCACCAATTTATCGATGA	660
Db	892	TCAATCTTTGA---CGGTGACCGGAAAAACGTTACAAAATGTACACAGCTTCACTGATGAA	948
QY	661	GGCAAGTGGATTCAGGTGATTAACCATACTTTTAAAGAACCCCTCACTATGTTGAAGATAG	720
Db	949	GGCACTACAGCTCAGCGGACACACCATTAAGGTAGAGATCTCTCACTAGGTAGAAAGATAA	1008
QY	721	GGCCATAAATATCTGTCTCTTTGAGCGAATACTGGAACAACGATGTTATCAAGCGAT	780
Db	1009	GGCCCAAAATACTTGTATTTTGAAACCAAACTGTGAACGTGAAGGCTACCAACGGCAA	1068
QY	781	CAGTTTCAATTAATTAAGCTTACTATGCGGGAAGTGACGTCCTTCTCCAGATGAAAAA	840
Db	1069	GAATCTTATTTTAAACAAAGCATACTATGTGCAAAAGCACATCTATTTCCGTCAGAAAGT	1128
QY	841	AATAACTGCTTCAAAAGTCTAAAAAACAAATGCTCTTTAGCGAATGTGCATTAGGC	900
Db	1129	CAAAAACCTTGTGCAAAAGCATTAATAAAAGCAGCGCTGATGATAGCAAAACGGCGCTCGGT	1188
QY	901	ATTGTTGAATGGCCGATGACTATATCAGTGAAGAAAGTGTATGAAACCATTAAGTCGATCA	960
Db	1189	ATGATTTAGCTTAAACGATGATTTACACTGTAAAAAAGATGAAACCGCTGATTCACATCT	1248
QY	961	AACACAGTAGCAGATGAAGTCGAGAGCGCCCAATATATTTAAATGAATTAATATGATAT	1020
Db	1249	AACACAGTAGCAGATGAATTAAGAGCGCGCAACGCTTTAAATTAAGACGCAATAGGTATC	1308
QY	1021	CTATTCACGATTCAGAAGATCCAAAATGACGAGTGAATTAACGACAAAGATGTT	1080
Db	1309	CTGTTCACTGACTCCCGCGGATCAAAAATGACGATTTGACGGGATTAAGCTCTAACGATATT	1368
QY	1081	TATATGTAAGGGCCCGGAGGCGACATCTTAAATGGCCCAACACACCGCATTAATGAACAT	1140
Db	1369	TACATAGCTTGGTATGTTTCTTAATCTTTTAAGTGGCCCATATCAACCGCTGTAAACAAACT	1428
QY	1141	GGACTTGTATTAAGACATGAATCTTGACCCCTGTGATCTCAACACACTTACTCTCATATGC	1200
Db	1429	GGCCTTGTGTTAAAAATGATCTTGATCTTAACGATGTAAACCTTTACTTACACACATTC	1488
QY	1201	GGTATCCCGCACCTGGAAGGTATATATGTGTACTACAAGTTATATGACGATAGAGGC	1260
Db	1489	GCTGTACTCTCAAGCGAAGGAAGCAATCTGTGATTCAACGCTATATGACAAACAGAGGA	1548



DB 1520 GCTGACCTCAAGCAAGAAACAATGCTGATTACAGCTATATGCAACAGAGCA 1579  
 QY 1261 TTCATCAGCAACATCACTCTCCCTGCGGACAGCTGGGCTTAATTAAGGCT 1320  
 DB 1580 TTCTACGACGACAAACATCAACCTTGGCGCAACCTCTCTGACATCAAGGCAAG 1639  
 QY 1321 GACACATCTGAGAGAGAAATAG 1343  
 DB 1640 AAAACATCTGTTGTCAGAGACAG 1662

RESULT 6  
 AAS68494  
 ID AAS68494 standard; cDNA; 1584 BP.  
 XX AAS68494;  
 AC  
 XX 13-FEB-2002 (first entry)  
 XX  
 XX DNA encoding novel human diagnostic protein #4298.  
 DE  
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 XX  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 XX (HSE-) HSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI: 2001-639362/73.  
 DR P-PSDB: ABG04307.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 1: SEQ ID No 4298; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1584 BP; 539 A; 358 C; 350 G; 337 T; 0 other;

Query Match 39.8%; Score 542.8; DB 23; Length 1584;  
 Best Local Similarity 65.9%; Pred. No. 2.4e-141;  
 Matches 875; Conservative 0; Mismatches 402; Indels 51; Gaps 4;

QY 16 TACAAAGAACTATGCTTTTCCCATATACAGCGCTGACATGCTAAATTCACAGA 75  
 DB 274 TACAAAGAAAGCTAGCGGCTCTCATATTAACGCCAGATATGCTCAGATCCCTAAA 333  
 QY 76 CACCAAAACAGTCCTCAATTTTAAGTTCATTCATGCAATCAATCAAAACACTT 135  
 DB 334 CAGCAGCAAAACAGAAATAATCCAGTGCCTCATTCATTCATCAACATTAATAATTT 393  
 QY 136 GATTGGGCAAAAGGCTATGATTAAGTCAGCACTTAATAGATTAGATGATGATAGC 195  
 DB 394 GAGTCTGCAAAAGACTT-----GAGTGTGGCAGACAGC 426  
 QY 196 TGGCCACTGCAAAACGCTGATGCTACGCGCAATTAATCATGATATACATGCTGCC 255  
 DB 427 TGGCCGCTGCAAAACGCTGACGAGACAGTACGTAATCAACGCTATACAGTGTGTT 486  
 QY 256 GCTTTAGAGAGTGCAGCAAAACAGATGATGCTACCTCATTTATTCATCAAAA 315  
 DB 487 GCTTTGGGGAGGCCGAAAGACCTGATGACACATCATCTACATGTTTATCAAAAG 546  
 QY 316 CTCGCTGATACATGATGACAGCTGAGAAATGCTGAGAGATATTGAAGATATGAT 375  
 DB 547 CTCGCGCAACACTCATATGACAGCTGAGAAATGCGGGCCGCTGTTTAAAGACGAT 606  
 QY 376 AATTTGTTCCAAATGATCCGATCTTAATATCAACACAGAGTGCAGTTCGCT 435  
 DB 607 AAGTTCGACGCCCAACGATCCGATCCTGAAGATCAGACGACAGATGTCGGTCTGCA 666  
 QY 436 ACTTAACCAAGATGCGCAACGCTGATCTATCTATACAGATTAATCTGATATCCGTA 495  
 DB 667 ACCTTTACATCTGAGGAAATATCCGTTATTTACACTGACTATTCGGTAAACATTA 725  
 QY 496 GATGCTGAGACCGCTGCTGTACCAAAATCATTTCACTGCTCAATCAACTATCCAG 555  
 DB 726 -----CGGCAACAAAGCCGTCAGACAGCCAGCTAAATGTCANAA 768  
 QY 556 CCGGATGCACTACCTTAAGTGCATGAGATATGATCATTAATCTGCTCTTGATGCGC 615  
 DB 769 TCTGAT---GACACACTCAAAATCAACGAGTGAAGATCAACAAACATTTTGA---C 822  
 QY 616 GGAGACGCTGACGTTATCAAAATATTCAGCAATTTATGATGAGGACGAGATTCA 675  
 DB 823 GGAGACGGAATAAATATTCAGATGAGATGATGATGATGATGATGATGATGATGAT 882  
 QY 676 GGTGATTAACATCTTTAAGAGACCTCATCTATGTTGAAGATTAAGGCCATTAATCTT 735  
 DB 883 GCGGACAAACATACCTGAGAGACCTCATCTATGTTGAAGACAAAGCCATTAATCTT 942  
 QY 736 GCTTTGAAGCAATAGTGAACACAGATGATGATGATGATGATGATGATGATGAT 795  
 DB 943 GATTTCGAAGCCAAACAGGGAACAGAAACGATACCAAGGAGAGATCTTTATTAC 1002  
 QY 796 AAGCTTCTATGCGGGAAGTACCTCTTCCAGATGAAATAAATAAATGCTTCA 855  
 DB 1003 AAGGCTACTACGCGGCGGACGAGATCTTCCCTAAGAAAGCCAAAGCTGAGAG 1062  
 QY 856 AGTCTTAATAAAACAATTTCTTTACCGAATGATGATGATGATGATGATGATGATG 915  
 DB 1063 AGCGTAAATAAAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1122  
 QY 916 GATGACTATACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 975  
 DB 1123 AATGATTAACATGTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1182  
 QY 976 GAAGTCGAAGCGCAATATATTTAATAATGATTAATAATGATTAATGATTAATGAT 1035  
 DB 1183 GAAATCGAGCGGCAATGTTTCAAAATGAAGGCAATGATGATGATGATGATGAT 1242  
 QY 1036 AGAGATCAAAATGACGATGATGATGATTAACGACAAAGATGTTATATCTAGGCGCC 1095

Db 1243 CGCGGTTCAAAAATGACGATGATGATTAACCTCAACAGATTTACATGCTGTTAT 1302  
 QY 1096 GGAGGCGACTCTCTTAATAGGCCACACACACCCGATTAATGAACCTGACTGTTATGAC 1155  
 Db 1303 GATCAAACTCTTTAAACCGCCCTTACAAAGCCGTAACAAAGGCGCTGCTGCA 1362  
 QY 1156 ATGAATCTGACCCCTGATCTACACACACTTACTGCTATGCGGATCCGACCT 1215  
 Db 1363 ATGGCTTGATCCAAAGATGTGACATTTACTTACTGCTGCACTGCGGACACC 1422  
 QY 1216 GAAGTAAATATGTTGTTACTACAAATATATGACGAATAGAGCTTCTATCCAGACAT 1275  
 Db 1423 AAAGGCAACAAATGCTGTTATCACAAGCTACATGACAAACAGAGCTCTTCCAGGATAAA 1482  
 QY 1276 CACTCTCAACCTGGGACAGAGCTGGGTTATATTAAGGCTGACACATCTGAGAGA 1335  
 Db 1483 AAGGCAACATTTGGCCCAAGCTTCTTAATGAACATCAAGGCAATTAACATCCGTTGTC 1542  
 QY 1336 GAAATAG 1343  
 Db 1543 AAAAACAG 1550  
 RESULT 7  
 AAS70417  
 ID AAS70417 standard; cDNA: 1935 BP.  
 AC AAS70417:  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #6221.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 PN W0200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-0508631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 DR WPI: 2001-639362/73.  
 DR P-PSDB: ABG06230.  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PS Claim 1; SEQ ID No 6221; 103pp: English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIR0  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.

SQ Sequence 1935 BP; 611 A; 468 C; 422 G; 434 T; 0 other:

Query Match 39.8%; Score 542.8; DB 23; Length 1935;

Best Local Similarity 65.9%; Pred. No. 2.6e-141;

Matches 875; Conservative 0; Mismatches 402; Indels 51; Gaps 4;

QY 16 TACAAGGAAGCATATGTTTGGCCATATATACAGCGCTGACATGTAATAATCCAGCA 75  
 Db 625 TACAAGGAAGCATATGTTTGGCCATATATACAGCGCTGACATGTAATAATCCAGCA 684  
 QY 76 CAACAAAACAGTCCATTTAAAGTGGCTCAATTCATGATGACGATCAATCAAAACAT 135  
 Db 685 CAGCAGCAAAACGAAATAATACCAAGTCCATTAATGATCAATCAATCAATCAATTA 744  
 QY 136 GATTGCGCAAAAGGCTATGATTAAGTCAGCAACTTAATAGATTAGATGATGGAATAGC 195  
 Db 745 GAGTGTGCAAAAAGGACTT-----GAGTGTGGGACAGC 777  
 QY 196 TGGCCACTGCAAAACGCTGATGTAAGTACGCGCAAAATATATGATGATACATGCTGCC 255  
 Db 778 TGGCCGCTGCAAAACGCTGACGAAACAGTACGATGATCAACGCGCTATGCTGTTT 837  
 QY 256 GCTTTAGCAGGTGACCCCAAAACAGTATGATCTCAGCTTCAATTTCTATCAAAA 315  
 Db 838 GCTCTGGGGGAAGCGCAAAAGACGCTGATGACATCAATCAATCTCAATTTATCAAAAG 897  
 QY 316 GTGCGTGAATACATGATGACAGCTGCAAAATCTGCAAGATATTTGAAGATATGAT 375  
 Db 898 GTGCGGCAACATCATGATGACAGCTGCAAAACGCGGCGCTGCTTTAAAGACAGCAT 957  
 Db 958 AAGTTCAGCGCAACAGATCGATCCCTGAAGATGACAGCAAGATGCGGTCTGCA 1017  
 QY 376 AAATTTGTTCCAAATGATCGTATCTTAATATCAAAACAGAGAGTGCTGAGTTCTGCT 435  
 Db 958 AAGTTCAGCGCAACAGATCGATCCCTGAAGATGACAGCAAGATGCGGTCTGCA 1017  
 QY 436 ACTTTAACCAAAAGATGCGCAAGCTGTTATTTCTATACAGTTACTACGATTAATCCGAA 495  
 Db 1018 ACCTTTACATCTGACGAAATAATCGTTTATTTCACTGACTATTCGGAACATTTA- 1076  
 QY 496 GATGTTGGAACCGGTGCTGATTAACCAATCAATTCACCTCAAGTAACTTATCCAG 555  
 Db 1077 -----CGGCAMCAAAAGCCTGACACAGCGCAGTAAATGTGTCAAAA 1119  
 QY 556 CCGATGAGCTACACTTAAGTGAAGTGAATCTGATCTGATCTGTTGATGAGC 615  
 Db 1120 TCTGAT---GACACACTCAAAATCAACGAGTGAAGTGAACCAAAAGCATTTTGA--C 1173  
 QY 616 GGAGACGCTACAGTTTATCAAAATATTTAGCAATTTATGATGAAGCAATGATGATTC 675  
 Db 1174 GGAGACGGAATAACATATCAACAGCTTCAAGATTTATGATGAAGCAATTTATCAATCC 1233  
 QY 676 GGTGATTAACATATCTTAAGAGACCCCTGATCTGTTGAAGTAAAGGCCATTAATATCT 735  
 Db 1234 GGCGAACACATACCTGATGAGACCCCTGATCTGTTGAAGCAAAAGGCCATTAATATCT 1293  
 QY 736 GCTTTGAAGGAATACTGCAACACAGATGTTATCAAGCGCATCTTCTTCAATAT 795  
 Db 1294 GTATTCAGGCAACACAGCGAAGAAACGATACCAAGGCAAGAAATCTTTATTAAC 1353  
 QY 796 AAGCTTACTATGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 855  
 Db 1354 AAGGTTACTATGCGGCGGCGGACAGCAATCTTCCGTAAGAAAGCAAGCAAGCTTCCAGCAG 1413







Db 918 ACCTTTACATCTGACGAAAAATCCGTTTATTCTACACTGACTATTCGCGTAACATTA- 860  
QY 496 GATGTTGACACCGGTGCTGTACCAAAATTCATTCAGTGGTCAATTAATTCACAG 555  
Db 859 -----CGGCAACAAAGCCTGACACACAGCGGACAGTAAATGTGTCAAAA 817  
QY 556 CCGATGACGCTACATTAAGTCGATGAGATATCTGATCAATTAATCTGTTGATGGC 615  
Db 816 TCTGAT---GACACACTCAAAATCAAGGAGTGGAGATCAACAAACATTTTGA---C 763  
QY 616 GAGACGCTACAGTTTATCAAAATTTTGAAGATTTATGATGAAGCAAGTGGATTGCA 675  
Db 762 GAGACGCGAAAAACATATCAAGACCTTCAGCAGTTTATGATGAAGCAATTTATACATCC 703  
QY 676 GGTGATATACATCTTTAGAGACCCCTACTATGTTTGAAGTAAAGGCCATTAATATCTT 735  
Db 702 GCGGCAACACATACCTGAGAGACCCCTACTATGTTGAGAGCAAAAGCCATTAATACCTT 643  
QY 736 GTCTTTGAGCGAATCTGAAACAACAGATGTTATCAAGCGCATCTTTCAATTAAT 795  
Db 642 GTATTGCAAGCAACACGGAACAGAAAAAGATACCAAGCGGAGAAATCTTTATTATAC 583  
QY 796 AAAGTTTACTATGGGGAAGTACGCTCTTCTCCAGATGAAAAAAATAAAGCTGTTCAA 855  
Db 582 AAAGGTTACTAGGGGCGGACGAACTCTTCGTTAAAGAACGCAAGCCTTCAGCAG 523  
QY 856 AGTCTAAACAAATGCTTCTTTAGCAGATGGTGGATTAGGCACTTTGAATTTGGCC 915  
Db 522 AGCGTTAAAAAACCGGATGCTGATGATGACGAACGGCCCTCGTATATATAGATTAAAT 463  
QY 916 GATGATCTATACAGTGAAGATGTTTGAACCATTTAGTCGATCAACACAGTACAGAT 975  
Db 462 AATGATTCACATGAAAAAAGTATGAAGCCGCTGATCACTTCAACACGATACGTAT 403  
QY 976 GAAGTCGAACCGGCAATATTTTAAATGAATATTAATGATCTATTCACGATTCATCA 1035  
Db 402 GAAATCGAGCGCGGAATGTTTCAAAATGAACGCAATGTTACTTGTTCACGTGATTC 343  
QY 1036 AGAGATCAAAATGACGAGTGAATGAACGAATTAACGAAGAATGTTTATATGCTAGCGCC 1095  
Db 342 CGCGTTCAAAAAATGACATGATGATGATTAATCAACGATATTTACATGCTTGGTTAT 283  
QY 1096 GAGGCGACTCTTAAATGCCCCACACACACCCGATTAATGAATGCACTTGTATTGAAAC 1155  
Db 282 GTATCAAACTCTTAAACCGCCCTTACCAAGCCGCTGAACAAACAGGCTTGTGCTGAA 223  
QY 1156 ATGATCTTGACCCCTGATCTCAGACACTTACTCTCATTTGGGATATCCGACCTT 1215  
Db 222 ATGGTCTTGAATCCAAACGATGTGACATTCATCTCACTTCGCGAGTCCGCAAGCC 163  
QY 1216 GAAGTATTAATGCTGATCTCAGCAAGTATATGAGCAATGAGGCTTCTATCCAGAACAT 1275  
Db 162 AAAGCAACATATGTTGATATCAACAGCTACATGACAAACAGAGCTTTTGGAGATTA 103  
QY 1276 CACTCTCAGCTCGCGGACAGCTTGGGTTAATATTAAGGCTTGACACATCTGGAGGA 1335  
Db 102 AAGCAACATTTGGCGCAAGCTTCTTAATGAACATCAAAAGCAATTAACATCCGTTGTC 43  
QY 1336 GAAATAG 1343  
Db 42 AAAAACAG 35

RESULT 10  
AAS69777 standard; cDNA: 2325 BP.

AC AAS69777;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #5581.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
OS WO200175067-A2.  
PN 11-OCT-2001.  
PD 30-MAR-2001; 2001WO-US08631.  
PE 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSE INC.  
PA Dmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
DR P-PSDB; ABG05590.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1: SEQ ID No 5581: 103pp: English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping; identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 2325 BP: 713 A; 593 C; 566 G; 453 T; 0 other:  
Query Match 39.8%; Score 542.8; DB 23; Length 2325;  
Best Local Similarity 65.9%; Pred. No. 2.8e-14;  
Matches 875; Conservative 0; Mismatches 402; Indels 51; Gaps 4;  
QY 16 TACAAGGAAGCACTAGTGTGGCCATTTACAGCGCGTGCACATGCTAAATTCAGGA 75  
Db 115 TACAAGGAAGCACTAGTGTGGCCATTTACAGCGCGTGCACATGCTAAATTCAGGA 174  
QY 76 CAACAAACAGCTCTCAATTAAGTGGCTCAATTCATGATGATGATCAAAACATTT 135  
Db 175 CAGCAGCAAAACGAAATATACCAAGTGGCTCAATTCATGATGATGATCAAAATTT 234  
QY 136 GATTGGCAAAAGGCTATGATTAAGTCAGCAACCTTAATAGATTAGATGATGGATAGC 195  
Db 235 GAGTCTGCAAAAGGACTT-----GATGTGGGACAGC 267  
QY 196 TGGCCACTGCAAAACGCTGATGATGCTGCGCAATTAATCATGATATCATGCTGCC 255  
Db 268 TGGCCGCTGCAAAACGCTGATGATGATGATGATGATGATGATGATGATGATGATG 327  
QY 256 GCTTTAGCAGGTGACCCCAAAACAGATGATGATGATGATGATGATGATGATGATGAT 315

```

Db      1384 AAAAAAG 1391
328 GCTCTGGGGAAGCCGAAAGACCTGATGACATCATCTTATCAAAAG 387
316 GTCCGTGATACATGATGACAGCTGGAAAAATGCTGGAAGATTTGAATATGAT 375
388 GTCCGGCACTCATCTAATCGAGCTGGAAAAACCGGCTTTTAAAGACGCGAT 447
376 AAATTTGTTCCAAATGATCCGTATCTTAAATATCAACACAGAGAGTGGCTGCT 435
448 AAGTTGACGCCAACGATCCGATCCTGAAAGATCAGACGAAGATGGTCCGGTTCTCA 507
436 ACTTTAACCAAGATGCGCAAGTCCGTTTATTTATACAGATTACTCAGTAATCTGAA 495
508 ACCTTTACATCTGACGAAAAATCCGTTTATCTACACTATTCGGTAAACATTA- 566
466 GATGTGGAACCGGTGCTGTATACCAATTCATTTCACTGCTCAGTAATCTATCCAG 555
567 -----CGCAAAACAAAGCCCTGACACAGCGCAGGTAAGTGTCAAA 609
556 CCGATGACGATACACTTAAAGTCAGTATGATCATTAATCTGCTTTGATGCG 615
610 TCTGAT---GACACACTCAAAATCAAGAGTGGAAAGTACAAACGATTTTTGA---C 663
616 GGAGACGTACAGTTTATCAAAATATTCAGCAATTTATGATGAAGCAAGTGAATTC 675
664 GGAGACGGAACCAATATCAGAACTGACAGATTTATGCAAGCAATATATCATCC 723
676 GGTGATACCACTATTAGAGACCTCATATGTTGAAGTAAGGGCATTAATATCTT 735
724 GCGCAACCACTAGCTGAGAGACCTCATCTGTTGAAGCAAGGCAATTAATCTT 783
736 GTCTTTGAAGCAATACTGAAACACAGATGTTTCAAGGCGATCAGTCTTCAATAT 795
784 GTATTGCAAGCAACACGGAACAGAAAGGATACAGAGGAGATCTTTATTATTAAC 843
796 AAAGCTTCTATGCGGGAATGACCTCTTCTTCCAGAAATGAAAAAATTAACCTGTTCA 855
844 AAAGGCTACTAGCGGCGGACGAGAACTTCTCCATAAGAAAGCAAGGCTTCAGAG 903
856 AGTCTAAAAAACAATGCTTCTTTAGCGAATGATGATTAAGCATTTGTAATGGCC 915
904 AGCGCTAAAAAACCGGATGCTGATGATGACGAGCGCCCTCGATATATAGTTAAT 963
916 GATGACTATACAGTGAAGATGTTATGAACCATTAAGTCGATCAACACAGATGACAGAT 975
984 AATGATTCACATTTAAAAAGTAAATGAAGCCGCTGATCATTCAACACGATACGAT 1023
976 GAACTGCAAGCGCCCAATATATTTAAATGAATTAATGATGATTTTACGAGATTCA 1035
1024 GAAATCGAGCGCGCAATGTTTCAAAATGAAGCGCAATGCTTCTGCTGATTTCA 1083
1036 AGAGATCCAAATGAGAGATGATGATTAAGCAAAAGATGTTATATGATAGGCGC 1095
1084 CGCGGTTAAAAATACATGATGATGATTTAACTAAACGATATTTACATGCTTGTTAT 1143
1096 GGAGCGACTCCTTAAATGAGCCACACACACCGGATTAATGAATGCACTGTTATGAAAC 1155
1144 GTATCAACCTCTTTAAACCGCCCTTAAACAGCGGCTGAACAAACAGGGCTTGCTGAA 1203
1156 ATGAATCTTGACCTGCTGATCTCAGACACTTACTCTCATTTGGGATATCCGACCTT 1215
1204 ATGGTCTTGATCCAAAGATGTGACATTTCACTTCTACTCTGCGAGTGGCGAACCC 1263
1216 GAAGTATATATGTTGTTACTCAAGTTATATGAGATGATGAGGTTTATCCAAACAT 1275
1284 AAAGCAACAATGTTGTTATCAAGCTACATGACAAACAGAGGCTTTCTGAGGATTA 1323
1276 CACTCTCACTGCGGACAGCTGGGTTATATTAATGAAGGCTGACACATCTGAGAGA 1335
1324 AAGGCAACATTTGGCCCAAGGCTTCTTAATGAACATCAAGGCAATAAACATCCGTTTC 1383
1336 GAAATATG 1343
1111 11

```

```

Db      1384 AAAAAAG 1391
RESULT 11
ID AAS75941 standard; cDNA: 2325 BP.
AA575941;
13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #11745.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Yang YT;
XX
DR WPI: 2001-639362/73.
XX P-PSDB; ABG11754.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
PS Claim 1; SEQ ID NO 11745; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX
CC Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2325 BP: 713 A; 593 C; 566 G; 453 T; 0 other:
XX
Query Match 39.8%; Score 542.8; DB 23; Length 2325;
Best Local Similarity 65.9%; Pred. No. 2.8e-141;
Matches 875; Conservative 0; Mismatches 402; Indels 51; Gaps 4;
OY 16 TACAAGGAAGACTAGTGTGGCCCATATTCACGCGGTGACATGCTAAATTCGAGGA 75
DB 115 TACAAGGAAGACTAGGCTGTCTCATATTACACGCGCATATATGCTGACAGATCCCTAAA 174

```

```

QY 76 CAACAAACAGCTCTCAATTTAAAGTCCTCAATTCATGCATCAGCAATCAAAAACATT 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 CAGCAGCAAAACGAAAGATACCAAGTGCCTCAATTCATGCATCAACGATTAATAATAT 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 136 GATTCGGCAAAAGGATATGATAGTCAGGCACTTAATAGATTAGATGTATGGATAGC 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 GAGTCTCGAAAAGGACTT-----GAGTGTGGGACAGC 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 196 TGGCCACTCAAAACGCTGATGCTACGCAATTCATGCATTCATCATGCTGCTCC 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 TGGCGGCTGCAAAACGCTGAGAGCAAGTACGTATTCACACGGCTATCAGCTGTGT 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 256 GCTTTAGCAGGTGACCAAAAACAGTGCATCTCCTCACTTATTCATCAAAA 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 GCTCTCGGGAAACCGGAAAGACGTGATGACATCATCTCATGTGTTTATCAAAAG 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 316 GTGGGTATTCATGATTCATGATGAGTGAAGAAATCTGGAAGATTTGAAGATATGAT 375
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 388 GTGGCGCAACTCAATCGACAGCTGGAAGAAACCGGCGCTGCTTTAAAGACAGCAT 447
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 376 AAATTTGTCCAATGATCCGTATCTTAATATCAAAACAGAGAGTGTGAGTTCTGCT 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 448 AAGTTCAGCGCAACGATCCGATCTGAAGATCAGACGCAAGATGTCGGTTCTGCA 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 436 ACTTTAAACAAAGATGGCCAAAGTCGTTATCTATACAGATTACTAGTAATCTGAA 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 508 ACCTTACATCTGACGAAATCCGTTTATCTACATGACTATTCGCGTAACATTA- 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 496 GATGCTGGAACCGGTGCTGTTAACCAATCTTTCACTGCTCAAGTAATCTTATCCAG 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 -----CGCAAAACAAACCTGACACAGCGGCGAGTAAAGTGTGCAAAA 609
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 556 CCGGATGACGCTCACTTAAAGTGCATGATCTATCTCATTAATCTGCTTTGATGGC 615
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 610 TCTGAT---GACACACTCAAAATCAACGAGTGAAGATTCACAAACGATTTTGA---C 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 616 GGAGAGCGTACAGTTTATCAAAATATTCAGCAATTCATGATGAAGGCAAGTGAATTTCA 675
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 664 GGAGACGGAATAATCATATCAGAACGTTTACAGCAATTCATGATGAAGGCAATTCATCATCC 723
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 676 GGTGATTAACCTACTTTAAGAGACCCCTCATCTATGTTGAAGTAAGGCGCATTAATCTT 735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 724 GGGCAGCAACATYAGCTGAGAGACCTCATACGTTGTAAGCAAAAGGCGCATTAATCTT 783
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 736 GTCTTTGAAGGAGTACTGGAACAGATGTTATCAAGCGCATCTTTCAATAT 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 784 GTATTGCAAGCAACAGGGAACAGAAACGATACCAAGCGAAGATCTTTATTTAAC 843
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 796 AAAGCTTACTATGCGGGAAGTGAAGTCTTTCAGAAATGAAGAAATTAACCTCTTCA 855
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 844 AAAGCTTACTATGCGGCGGACGAACTTCTTCAAAAGCAAGCAAGCAAGCTTCAGCAG 903
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 AGTCTTAAAAACAATTTGCTTCTTTAGCGAATGCTGATTAGGATTTGTAATGGCC 915
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 904 AGCGCTAAAAACCGGATGCTGAGTGAAGCAAGCGCGCTCGGTATCATAGAGTTAAT 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 916 GATGACTATACAGTAAAGTGTATGAACCACTTATGCGCAAAACAGTACAGTACAGAT 975
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 AATGATTACACATTTGAAAAAGTATAGAGCGCTGATCACTTAAACACGCTATACAT 1023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 976 GAAGTCAGACGCGCAATATATTTAAATGAATATAATGATCTATATTCAGCAGATTTCA 1035
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 GAATTCGAGCGCGGCAATGTTTCAAAATGAAGCAAAATGATCTGTTTCAGATTTCA 1083
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1036 AGAGATCCAAATAGAGAGATGATGAATTAAGCAAAAGTGTATATCTAGGGCC 1095
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 CGCGGTTCAAAATAGAGATGATGATTAATCAACGATTTATACCTCTGTTAT 1143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1096 GGAGGCACTCTTAATATGCGCCACACACAGCGATTAATGAATGCACTGATTTGATAC 1155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 GTATCAAACTTTTAACCGGCTTACAGAGCGCTGAACAAACAGGCGTTGTCTGCA 1203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1156 ATGAATCTTGACCTGCTGATCTCACACACACTTACTCTCATTCGCGGTATCCGACACT 1215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1204 ATGGCTTTGATCCAAACGATGTGACATTCATCTACTCTGCGAGTCCCAAGCC 1263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1216 GAAGTATATATGCTGCTACTCACAAGTATATATGACGATTAAGGCTTATCAGAACAT 1275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1264 AAAGCAACAATGTGTTATCACAAGCTTACATGACAAACAGAGCTTCTGAGGATAA 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1276 CACTCTCACCTGCGGACACACTTGGGTTATATTAAGGCTGACACATCTGAGAGA 1335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1324 AAGGCAACATTTGCCCAAGCTTTTATGAAATCAAAAGCAATAAACATCCGTTGTC 1383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1336 GAAATATAG 1343
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1384 AAAAACAG 1391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AA010498
ID AA010498 standard; DNA; 2351 BP.
XX
AC AA010498;
XX
DT 17-APR-1991 (first entry)
XX
DE B. amylioliquefaciens sacB[Bamp] fragment.
XX
KW levensucrase; sucrose regulation; Bacillus; SacB[Bamp]; ss.
XX
OS Bacillus amylioliquefaciens.
XX
FH Key Location/Qualifiers
FT sig_peptide 863..950
FT mat_peptide 951..2279
FT terminator 2291..2327
FT misc_feature 706..764
FT /tag= c
FT /note= "rho-independent"
FT /tag= d
FT /function= putative regulatory region
XX
PN WO9100913-A.
XX
PD 24-JAN-1991.
XX
PE 20-JUN-1990; 90WO-US03348.
XX
PR 07-JUL-1989; 89US-0376474.
XX
PA (DUPO ) DU POWT DE NEMOURS CO.
XX
PI Nagarajan V, Tang LB;
XX
DR MPI; 1991-051337/07.
XX
DR P-PSDB; AAR10671.
XX
PT Sucrose regulatable expression vector which can replicate -
PT derived from non-Bacillus subtilis Bacillus species for
PT expression in many bacterial species
XX
PS Example; Fig 3; 44pp; English.
XX
CC The levensucrase-encoding sequence was isolated from Bacillus
CC amylioliquefaciens lambda ZAP phage plaques. After three successive
CC screening and re-cloning steps, (using probes based on the sequence
CC of B. subtilis sacB), all the plaques were found to hybridise to the
CC probes. Clones 2A and 2C containing the putative scab[Bamp] gene
CC sequences were converted to Bluescript to give plasmids pBS300 and
CC pBS301, respectively. DNA was isolated from the plasmids and
CC digested with EcoRI in preparation for a Southern transfer. The
CC same sacB[Bsu] probes were used in the Southern hybridisation and

```

CC both hybridised only with a 1.5kb fragment. Sequence analysis of  
CC this fragment showed it to be the sacB[bamp] gene. The sucrose  
CC regulatory region can be incorporated into stable vectors to  
CC control expression of heterologous genes in transformed B.subtilis.  
CC See also AMQ10497-7.

XX Sequence 2351 BP; 739 A; 524 C; 506 G; 582 T; 0 other;

Query Match 39.8%; Score 542.8; DB 12; Length 2351;  
Best Local Similarity 65.9%; Pred. No. 2.9e-141;  
Matches 875; Conservative 0; Mismatches 402; Indels 51; Gaps 4;

```

OY 16 TACAAGAAAGATGTTTGGCCATATTCACGCGGTGACATCTTAAAAATCCAGCA 75
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 972 TACAAGAAAGATGTTTGGCCATATTCACGCGGTGACATCTTAAAAATCCAGCA 1031
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 76 CAACAAAACAGTCCCAATTTAAAGTGGCTCAATTCATGCATGCAATCAAAAATT 135
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1032 CAGCGCAAAAAGAAAATACCAAGTGCCTCAATTCATGCATCAACATTTAAATATT 1091
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 136 GATTGGCAAAAAGGTTATAGTCAGGCAACTTAATAGATTATGATGATGATAGC 195
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1092 GAGTGTGCAAAAAGACTT-----GATGTGGGACAGC 1124
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 196 TGGCCACTGCAAAACGCTGATGCTACTCGGCAAAATATCATGATATCATGCTGCC 255
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1125 TGGCCGCTGCAAAACGCTGAGGAGACAGTATGATCAACGCGCTATCACGTTGT 1184
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 256 GCTTATGAGGATGACCAAAAACAGTATGATGATGATGATGATGATGATGATGAT 315
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1185 GCTTATGAGGATGACCAAAAACAGTATGATGATGATGATGATGATGATGATGAT 1244
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 316 GTCGCTGATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1245 GTCGCGCACTCACTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1304
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 376 AAATTTGTTCCAAATGATCCGATATTTAAATATCAACACAGAGTGTGCTGCT 435
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1305 AAGTTTCAGCGCCACAGATCCGATCCCTGAAAGATCAGACGCAAGATGTCGCTGCA 1364
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 436 ACTTAAACCAAGATGAGGCAAGTCCGTTATCTTATACAGATGATGATGATGATGAT 495
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1365 ACCTTATACATCTGACGGAATAATCCGTTTATTTACACGACTATTCGGTAAACATT 1423
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 496 GATGCTGAGACCGGTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 555
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1424 -----CGGCAACCAAGCCGTCGACACAGCGCAGTAAATGTGCAAA 1466
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 556 CCGGATGCACTGATCACTTAAAGTGCATGATGATGATGATGATGATGATGATGAT 615
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1467 TCTGTAT---GACACACTCAAAATCAACGAGTGAAGATCAACAAAACGATTTTGA---C 1520
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 616 GGAGACGGTACAGTTATCAAAAATATTCAGCAATTTATGATGAGGCAAGTGGATTTCA 675
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1521 GGAGACGGGAAAACATATCAGACCTTACAGCTTATGATGATGATGATGATGATGAT 1580
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 676 GGTGATTAACATCTTAAAGACCCCTCACTATGTTGAAGATTAAGGCCATTAATATCT 735
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1581 GGCGACACCAATACCTGAGAGACCCCTCACTATGTTGAAGAACAAAGGCCATTAATACCT 1640
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 736 GTCTTTGAAGCAATCTGGAACACACAGATGTTATCAAGGCGATCCTTTCAATAT 795
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1641 GTATTCGAGCCACCAACGAGACAGAAAACGATACCAAGAGAAATCTTATTATTAAC 1700
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 796 AAAGCTTACTATGAGGGAAGTGAAGTCTTCTCCAGATGAAGAAAATTAACCTGTTCAA 855
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1701 AAAGCTTACTATGAGGGAAGTGAAGTCTTCTCCAGATGAAGAACAGCAAGGTTAGCAG 1760
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 856 AGTCTTAAAAAACAATGCTTCTTTAGCGATGATGATGATGATGATGATGATGATGAT 915
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1761 AGCGCTTAAAAAAGCGGATGCTGATGATGATGATGATGATGATGATGATGATGAT 1820
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 916 GATGACTATACAGTGAAGAAAGTGTATGAAACCAATTAGTGGCATCAACACAGTACAGAT 975

```

```

Db 1821 AATGATTTACATGTAAGAAAAATGATGAGCCGCTGATCACTTCAACACGTAATCAT 1880
OY 976 GAAGCGCAAGCCGCAATATATTTAAATAATGATGATGATGATGATGATGATGATGAT 1035
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1881 GAAATCGAGCGCGGAATGTTTCAAAATGAAACGCAATATGATGATGATGATGATGAT 1940
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1036 AGAGATCCAAATGACAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1941 CCGGTTTCAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2000
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1096 GGAGCGACCTCTTAATATGSCCACAACACCCGATTAATGAAGTGAAGTGAAGTGAAG 1155
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2001 GTATCAAACTCTTTAAACCGCGCTTACAGCGCTGAAACAAACAGGCTTGTCTGCA 2060
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1156 ATGATCTTGAACCTCGCATCTGACACACACTTACTCTGCTGCGGATATCCGACCC 1215
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2061 ATGGCTTGTATCCAAACGATGATGATGATGATGATGATGATGATGATGATGATGAT 2120
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1216 GAAGTAAATATGTTGTTACTCACAAGTATATGACGATGATGATGATGATGATGAT 1275
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2121 AAAGCAACATATGTTGTTATCACAAGCTATGATGATGATGATGATGATGATGATGAT 2180
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1276 CACTCTACCTGCGGACACACTTGGGTTAATTAAGGCTGACACATCTGAGAGA 1335
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2181 AAGGCACATTTGGGCCAAGCTCTTAAATGAACATCAAAAGCAATAAACATCCGTTGTC 2240
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1336 GAAATATG 1343
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2241 AAAACAG 2248

```

## RESULT 13

AAS69133  
ID AAS69133 standard; cDNA: 2370 BP.

AAS69133;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #4937.

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

WO200175067-A2.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Dmanac RT, Liu C, Tang YT;

WPI: 2001-639362/73.

P-PSDB: ABG04946.

New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity.

Claim 1; SEQ ID No 4937; 103bp; English.

The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPRO  
 CC at ftp.wipro.int/pub/published\_pct\_sequences.

Sequence 2370 BP; 743 A; 535 C; 544 G; 548 T; 0 other;

Query Match 39.8%; Score 542.8; DB 23; Length 2370;

Best Local Similarity 65.9%; Pred. No. 2.9e-141; Indels 51; Gaps 4;

Matches 873; Conservative 0; Mismatches 402; Indels 51; Gaps 4;

QY 16 TACAAGAGAGACTGTTTGGCCATTTACAGCGCTGACATGCTAAATTCAGGA 75  
 DB 1060 TACAAGAGAGAGACTGTTTGGCCATTTACAGCGCTGACATGCTAAATTCAGGA 1119  
 QY 76 CACAAAGAGCTCTCAATTTAAAGTGGCTCAATTCAGTACAGCAATCAAAACAT 135  
 DB 1120 CAGAGCAAAAG 1179  
 QY 136 GATTCGGCAAAAG 195  
 DB 1180 GAGTCTGCAAAAG 1212  
 QY 196 TGGCCACTCAAAAG 255  
 DB 1213 TGGCCGCTCAAAAG 1272  
 QY 256 GCTTACAGAGTGAAG 315  
 DB 1273 GCTTGGGAG 1332  
 QY 316 GTCGATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375  
 DB 1333 GTCGATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1392  
 QY 376 AAATTTGTTCCAAATGATCCGATCTTAATATCAAAACAGAGAGAGAGAGAGAG 435  
 DB 1393 AAGTTGAG 1452  
 QY 436 ACTTTAACAAG 495  
 DB 1453 ACCTTTACATGCTGAG 1511  
 QY 496 GATGATGAG 555  
 DB 1512 -----CGGCAAAAG 1554  
 QY 556 CCGGATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 615  
 DB 1555 TCTGAT---GACACACTCAAAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608  
 QY 616 GGAAG 675  
 DB 1609 GGAAG 1668  
 QY 676 GGTGATACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 735  
 DB 1669 GGGGCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1728

QY 736 GCTTTGAAG 795  
 DB 1729 GATTTGAG 1788  
 QY 796 AAAGCTTACTATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 855  
 DB 1789 AAAGGCTACTGCGGAG 1848  
 QY 856 AGCTTAAAG 915  
 DB 1849 AGCGTTAAAG 1908  
 QY 916 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 975  
 DB 1909 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1968  
 QY 976 GAAGTGAAG 1035  
 DB 1969 GAATGAG 2028  
 QY 1036 AGAGATGCAAAAG 1095  
 DB 2029 CGCGTTCAAAAG 2088  
 QY 1096 GAG 1155  
 DB 2089 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2148  
 QY 1156 ATGAATCTTGAAG 1215  
 DB 2149 ATGGTCTTGAAG 2208  
 QY 1216 GAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1275  
 DB 2209 AAAGGCAAAAG 2268  
 QY 1276 CACTCTCACTGCGGAG 1335  
 DB 2269 AAGGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2328  
 QY 1336 GAAATAG 1343  
 DB 2329 AAAAAGAG 2336

RESULT 14  
 AAS92594  
 ID AAS92594 standard; cDNA: 2370 BP.  
 XX  
 AC AAS92594;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #28398.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 DT drmanac RT, Liu C, Tang YT;  
 XX









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 02:52:36 ; Search time 63.317 Seconds  
(without alignments)  
6611.951 Million cell updates/sec

Title: US-09-986-682B-4

Perfect score: 1365

Sequence: 1 ATGAACAGCGGGGAGCTACAA.....CCGACAGAGACATTCACA 1365

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata1/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata1/lna/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata1/lna/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata1/lna/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1365	100.0	1365	2	US-08-870-827-4
2	1365	100.0	1365	4	US-09-317-179-4
3	1365	100.0	2408	4	US-08-870-827-5
4	1365	100.0	2408	4	US-09-317-179-5
5	577	42.3	10317	3	US-09-058-746-1
6	577	42.3	10317	3	US-09-438-142-1
7	539.6	39.5	3305	3	US-09-068-043-1
8	43.4	3.2	7218	1	US-08-232-463-14
9	39.8	2.9	1662	1	US-08-671-947-1
10	36.6	2.7	7101	2	US-08-480-604-9
11	36.6	2.7	7101	2	US-08-405-496A-9
12	36.6	2.7	7101	4	US-08-915-136-9
13	36.6	2.7	7101	4	US-08-857-310-9
14	36.4	2.7	3095	6	5231168-1
15	35.4	2.6	849	4	US-09-648-520E-48
16	35.4	2.6	3317	4	US-09-193-562D-1
17	34.8	2.5	62	3	US-09-068-043-5
18	34.2	2.5	3360	1	US-07-712-833A-1
19	34	2.5	1467	4	US-09-134-001C-1409
20	34	2.5	2817	5	PCT-US93-05944-1
21	34	2.5	3356	4	US-09-379-523-4
22	33.8	2.5	3279	4	US-08-446-137B-1
23	33.6	2.5	2408	2	US-08-870-827-5
24	33.6	2.5	2408	4	US-09-317-179-5
25	33.6	2.5	3713	1	US-08-100-709-1
26	33.6	2.5	3713	1	US-08-176-865-1
27	33.6	2.5	3713	1	US-08-474-038-1

28	33.6	2.5	3713	2	US-08-779-046-1	Sequence 1, Appl
29	33.6	2.5	3713	2	US-08-881-340-1	Sequence 1, Appl
30	33.4	2.4	1566	4	US-09-134-001C-2802	Sequence 2802, Ap
31	33.2	2.4	900	4	US-08-879-098-1	Sequence 1, Appl
32	33.2	2.4	900	4	US-09-631-548-1	Sequence 1, Appl
33	33.2	2.4	3891	1	US-08-480-604A-27	Sequence 27, Appl
34	33.2	2.4	3891	2	US-08-405-496A-27	Sequence 27, Appl
35	33.2	2.4	3891	4	US-08-915-136-27	Sequence 27, Appl
36	33	2.4	699	4	US-08-936-165A-233	Sequence 233, App
37	33	2.4	2047	4	US-08-836-261A-1	Sequence 1, Appl
38	33	2.4	4163	4	US-09-004-838-70	Sequence 70, Appl
39	33	2.4	5820	4	US-09-029-213B-7	Sequence 1641, Ap
40	32.8	2.4	630	4	US-09-134-001C-1641	Sequence 1, Appl
41	32.8	2.4	2600	3	US-08-988-251-1	Sequence 1, Appl
42	32.8	2.4	2600	3	US-09-386-048-1	Sequence 1, Appl
43	32.8	2.4	3280	1	US-08-259-000-4	Sequence 4, Appl
44	32.8	2.4	3280	1	US-08-729-767-6	Sequence 6, Appl
45	32.8	2.4	8195	4	US-08-961-527-94	Sequence 94, Appl

## ALIGNMENTS

```
RESULT 1
US-08-870-827-4
: Sequence 4, Application US/08870827
: Patent No. 5962297
: GENERAL INFORMATION:
: APPLICANT: Tsusaki et al.
: TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
: TITLE OF INVENTION: ACTIVITY
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/870,827
: FILING DATE: 06-JUN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 170,630/1996
: FILING DATE: 10-JUN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: TSUSAKI-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1365 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: US-08-870-827-4
:
: Query Match 100.0%; Score 1365; DB 2; Length 1365;
: Best Local Similarity 100.0%; Pred. No. 0;
: Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 1 ATGAACAGCGGGGAGCTACAAAGAACTATGTTTGGCCCATATTTACAGCGCTGACATG 60
: Db 1 ATGAACAGCGGGGAGCTACAAAGAACTATGTTTGGCCCATATTTACAGCGCTGACATG 60
```

```
QY 61 CTAATAATTCAGACACAAGAGTCCCAATTTAAAGTCCCAATTCATGATCA 120
DB 61 CTAATAATTCAGACACAAGAGTCCCAATTTAAAGTCCCAATTCATGATCA 120
QY 121 GCAATCAAAAACATTTGATGGGCAAAAAGGATGATGAAGTCAAGCACTTAATGATTA 180
DB 121 GCAATCAAAAACATTTGATGGGCAAAAAGGATGATGAAGTCAAGCACTTAATGATTA 180
QY 181 GATGATGGGATGATGGGCACTGCAAAAGCTGATGGTACTGGGCAAAATTCATGATGA 240
DB 181 GATGATGGGATGATGGGCACTGCAAAAGCTGATGGTACTGGGCAAAATTCATGATGA 240
QY 241 TATGACATGCTCTCGCTTTAGACAGTGATGACCAAAAACAGTGAATGATGATGATGAT 300
DB 241 TATGACATGCTCTCGCTTTAGACAGTGATGACCAAAAACAGTGAATGATGATGATGAT 300
QY 301 TTATTTCTATCAAAAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 TTATTTCTATCAAAAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 TTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 TTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 TGGTCAAGTCTGCTACTTTTAAACCAAGATGACCAAGTCCGTTTATTTCTATGATGAT 480
DB 421 TGGTCAAGTCTGCTACTTTTAAACCAAGATGACCAAGTCCGTTTATTTCTATGATGAT 480
QY 481 TCAGGATATCTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
DB 481 TCAGGATATCTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 541 GTAAAGTATCCGACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 541 GTAAAGTATCCGACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 601 TCTGCTTTGATGGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 601 TCTGCTTTGATGGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 GGCAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 661 GGCAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 721 GGCATTAATATCTGCTTTGAGGAGGATGATGATGATGATGATGATGATGATGATGAT 780
DB 721 GGCATTAATATCTGCTTTGAGGAGGATGATGATGATGATGATGATGATGATGATGAT 780
QY 781 CAGTCTTTCAATTAATAAGCTTACTATGCGGAGGATGATGATGATGATGATGATGAT 840
DB 781 CAGTCTTTCAATTAATAAGCTTACTATGCGGAGGATGATGATGATGATGATGATGAT 840
QY 841 AATAAGCTGCTTCAAAATCTTAAATAAGCTTACTATGCGGAGGATGATGATGATGAT 900
DB 841 AATAAGCTGCTTCAAAATCTTAAATAAGCTTACTATGCGGAGGATGATGATGATGAT 900
QY 901 ATTGTTGAATTTGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 901 ATTGTTGAATTTGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 AACACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1020
DB 961 AACACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1020
QY 1021 CTATTCAGGATTTAAAGAGTCCAAATATGATGATGATGATGATGATGATGATGATGAT 1080
DB 1021 CTATTCAGGATTTAAAGAGTCCAAATATGATGATGATGATGATGATGATGATGATGAT 1080
QY 1081 TATATGCTAGGAGCCGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 TATATGCTAGGAGCCGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
```

```
QY 1141 GGACTGTATTTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 GGACTGTATTTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 GGTATCCCGGACCTGGAAGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 GGTATCCCGGACCTGGAAGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 TTCTATCCAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 TTCTATCCAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 GACACATCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1365
DB 1321 GACACATCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1365

RESULT 2
US-09-317-179-4
: Sequence 4, Application US/09317179
: Patent No. 6383769
: GENERAL INFORMATION:
: APPLICANT: Tsusaki et al.
: TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/317,179
: FILING DATE: 24-May-1999
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/870,827
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: TSUSAKI-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1365 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: Linear
: SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-317-179-4
Query Match 100.0%; Score 1365; DB 4; Length 1365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAACAGCGGGGAGTCAAGAGGAGTATGTTTGGCCATATTTACAGCGCGTGACATG 60
DB 1 ATGAACAGCGGGGAGTCAAGAGGAGTATGTTTGGCCATATTTACAGCGCGTGACATG 60
QY 61 CTAATAATTCAGACACAAGAGTCCCAATTTAAAGTCCCAATTCATGATCA 120
DB 61 CTAATAATTCAGACACAAGAGTCCCAATTTAAAGTCCCAATTCATGATCA 120
```

QY	121	GCATTCACAAAACATTTGATTGGCGCAAAAGGGTGTGAATGATCGGCAACCTTAATTAATTTA	180
Db	121	GCATTCACAAAACATTTGATTGGCGCAAAAGGGTGTGAATGATCGGCAACCTTAATTAATTTA	180
QY	181	GATGATAGGGAGTAGCTGGCCACTGCGAAAAAGCGTGTAGTGTACTGCGCGCAAAATTAATGATGA	240
Db	181	GATGATAGGGAGTAGCTGGCCACTGCGAAAAAGCGTGTAGTGTACTGCGCGCAAAATTAATGATGA	240
QY	241	TATCACAATCGTCTCCGCTTTAGCAGGTGACCCAAAAAACAGTGAATGATCTCCACTTCAT	300
Db	241	TATCACAATCGTCTCCGCTTTAGCAGGTGACCCAAAAAACAGTGAATGATCTCCACTTCAT	300
QY	301	TTATTCCTATCAAAAAGTGGGTGATACATCGATGTGACAGTGGAAAAATTCCTGGAGAGTA	360
Db	301	TTATTCCTATCAAAAAGTGGGTGATACATCGATGTGACAGTGGAAAAATTCCTGGAGAGTA	360
QY	361	TTTGAAGATATGAGATAAATTTGTTCCAAATGATCCGTAATCTTAAATATCAACACAGGAG	420
Db	361	TTTGAAGATATGAGATAAATTTGTTCCAAATGATCCGTAATCTTAAATATCAACACAGGAG	420
QY	421	TGGTCAAGTTCTGCTACTTTAAACAAGATGGCCAAAGTCCGTTATTCATATACAGATTAC	480
Db	421	TGGTCAAGTTCTGCTACTTTAAACAAGATGGCCAAAGTCCGTTATTCATATACAGATTAC	480
QY	481	TCAGGTAATCCGAGATAGTGGTGAACCGGTGCTGTGTAAACCAATCATTTCAACTCTCTCA	540
Db	481	TCAGGTAATCCGAGATAGTGGTGAACCGGTGCTGTGTAAACCAATCATTTCAACTCTCTCA	540
QY	541	GTTAAACTTATCCACACCGGATGCACACTACCTTAAAGTCGATGGAGTATCTGATATATA	600
Db	541	GTTAAACTTATCCACACCGGATGCACACTACCTTAAAGTCGATGGAGTATCTGATATATA	600
QY	601	TCGTCTTTGATGGCGGAGACGGTACAGTTTATCAAAAATATTACAGCAATTTATCGATGA	660
Db	601	TCGTCTTTGATGGCGGAGACGGTACAGTTTATCAAAAATATTACAGCAATTTATCGATGA	660
QY	661	GGCAGTGGATTTTCAGGTGATPACCATACTTTAAAGAGACCCCTCATATGTTGMAATGAAG	720
Db	661	GGCAGTGGATTTTCAGGTGATPACCATACTTTAAAGAGACCCCTCATATGTTGMAATGAAG	720
QY	721	GGCCCTTAATATCTGTCTTTTGAACGGAATACGTGAAACAACGATGGTATCAAGCGCAT	780
Db	721	GGCCCTTAATATCTGTCTTTTGAACGGAATACGTGAAACAACGATGGTATCAAGCGCAT	780
QY	781	CAGTCTTCAATATATTAAGCTTACTATGCGGGAATGACGTCCTTTCCAGAGATGAAAA	840
Db	781	CAGTCTTCAATATATTAAGCTTACTATGCGGGAATGACGTCCTTTCCAGAGATGAAAA	840
QY	841	AATAAAGCTGCTTCAAGAGTCCTTAAACCAAAATGCTTTTACGAAATGCTATTAAGC	900
Db	841	AATAAAGCTGCTTCAAGAGTCCTTAAACCAAAATGCTTTTACGAAATGCTATTAAGC	900
QY	901	ATTGTTGAATTTGGCCGATGACTATACAGTGAATAATGTATGAAACCAATTAAGTCCATCA	960
Db	901	ATTGTTGAATTTGGCCGATGACTATACAGTGAATAATGTATGAAACCAATTAAGTCCATCA	960
QY	961	AACACAGTAGCAGATGAAGTGAAGCGCCCAATATATTTAAATGAATAAATTAAGTGAAT	1020
Db	961	AACACAGTAGCAGATGAAGTGAAGCGCCCAATATATTTAAATGAATAAATTAAGTGAAT	1020
QY	1021	CTATTCAAGGATTCAGAAGGATCCAAAATGACGAGTGAATGCAATTAACGACAAAGATGT	1080
Db	1021	CTATTCAAGGATTCAGAAGGATCCAAAATGACGAGTGAATGCAATTAACGACAAAGATGT	1080
QY	1081	TATATGCTAGGGCCCGAGGCGACTCTTAAATGGCCCAACACCCGATTAATTAAGTAACT	1140
Db	1081	TATATGCTAGGGCCCGAGGCGACTCTTAAATGGCCCAACACCCGATTAATTAAGTAACT	1140
QY	1141	GGACTTTGATTGAAACATGATCTTGAACCTGCTGATCTCACACACACTTACTCTGATTCG	1200
Db	1141	GGACTTTGATTGAAACATGATCTTGAACCTGCTGATCTCACACACACTTACTCTGATTCG	1200
QY	1201	GGTATCCCGCACCTCGAAGGTATATATGTGTACTCACAGTTATATACGAATATGAGGC	1260

Accession	Sequence	Length
Dd	1201 GGTATCCCGCACCCCTGAAGGTAATAATGTTGGTACTCACAAAGTTATATGACCAATATAGAGGC	1260
Oy	1261 TTCTATCCAGAACATCACTCTCACCTGGCGGAGCAAGCTTTGGGGTTAAATATTTAAAGGCTCT	1320
Dd	1261 TTCTATCCAGAACATCACTCTCACCTGGCGGAGCAAGCTTTGGGGTTAAATATTTAAAGGCTCT	1320
Oy	1321 GACCAATCTGGAGAGAAATAGTTCCGGCAAGCAAGCAATTTCCCA	1365
Dd	1321 GACCAATCTGGAGAGAAATAGTTCCGGCAAGCAAGCAATTTCCCA	1365

```

1      RESULT 3
2      US-08-870-827-5
3      Sequence 5, Application US/08870827
4      Patent No. 5962297
5      GENERAL INFORMATION:
6      APPLICANT: Tsusaki et al.
7      TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOPURANOSIDASE
8      TITLE OF INVENTION: ACTIVITY
9      NUMBER OF SEQUENCES: 6
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: BROWDY AND NEWMARK
12     STREET: 419 Seventh Street, N.W., Suite 300
13     CITY: Washington
14     STATE: D.C.
15     COUNTRY: USA
16     ZIP: 20004
17     COMPUTER READABLE FORM:
18     MEDIUM TYPE: Floppy disk
19     COMPUTER: IBM PC compatible
20     OPERATING SYSTEM: PC-DOS/MS-DOS
21     SOFTWARE: PatentIn Release #1.0, Version #1.30
22     CURRENT APPLICATION DATA:
23     APPLICATION NUMBER: US/08/870,827
24     FILING DATE: 06-JUN-1997
25     CLASSIFICATION: 435
26     PRIOR APPLICATION DATA:
27     APPLICATION NUMBER: JP 170,630/1996
28     FILING DATE: 10-JUN-1996
29     ATTORNEY/AGENT INFORMATION:
30     NAME: BROWDY, Roger L.
31     REGISTRATION NUMBER: 25,618
32     REFERENCE/DOCKET NUMBER: TSUSAKI-2
33     TELECOMMUNICATION INFORMATION:
34     TELEPHONE: 202-628-5197
35     TELEFAX: 202-737-3528
36     INFORMATION FOR SEQ ID NO: 5:
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 2408 base pairs
39     TYPE: nucleic acid
40     STRANDEDNESS: double
41     TOPOLOGY: linear
42     MOLECULE TYPE: genomic DNA
43     ORIGINAL SOURCE:
44     ORGANISM: Bacillus sp.
45     INDIVIDUAL ISOLATE: V230 (FERM BP-5054)
46     FEATURE:
47     NAME/KEY: 5'UTR
48     LOCATION: 1..360
49     IDENTIFICATION METHOD: E
50     NAME/KEY: signal peptide
51     LOCATION: 361..456
52     IDENTIFICATION METHOD: S
53     NAME/KEY: mat peptide
54     LOCATION: 457..1821
55     IDENTIFICATION METHOD: S
56     NAME/KEY: 3'UTR
57     LOCATION: 1822..2408
58     IDENTIFICATION METHOD: E
59     US-08-870-827-5

```

Query Match	100.0%;	Score 1365;	DB 2;	Length 2408;
-------------	---------	-------------	-------	--------------

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 ATGAAACGCGGGGACTACAGAGACTATGTTGGCCATTTACAGCGGCTGACATG 60
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 457 ATGAAACGCGGGGACTACAGAGACTATGTTGGCCATTTACAGCGGCTGACATG 516
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 61 CTAAAAATTCAGAGACAACAAGAGTCTCAATTTAAAGTCCATTAATGATCATCA 120
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 517 CTAAAAATTCAGAGACAACAAGAGTCTCAATTTAAAGTCCATTAATGATCATCA 576
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 121 GCAATCAAAAACATTTGATGCGCAAAAAGGTATGATTAAGTACAGCACTTAATGATTA 180
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 577 GCAATCAAAAACATTTGATGCGCAAAAAGGTATGATTAAGTACAGCACTTAATGATTA 636
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 181 GATGATGGATGATGCGCACTGCAAAAGCTGATGCTAGCGGCAATTAATCATGGA 240
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 637 GATGATGGATGATGCGCACTGCAAAAGCTGATGCTAGCGGCAATTAATCATGGA 696
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 241 TATGACATCGTCTCCGCTTTAGCAGGTGACCCAAAACAGTGAATGATCTCACTTCAT 300
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 697 TATGACATCGTCTCCGCTTTAGCAGGTGACCCAAAACAGTGAATGATCTCACTTCAT 756
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 301 TTATTCATCAAAAAGTGGTGTATCATGATGACAGCTGGAATAATGCTGGAAGACTA 360
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 757 TTATTCATCAAAAAGTGGTGTATCATGATGACAGCTGGAATAATGCTGGAAGACTA 816
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 361 TTGGAATATGATGATTAATTTGTCCAAAATGATCGGTATCTTAATATCAAAACAGGAG 420
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 817 TTGGAATATGATGATTAATTTGTCCAAAATGATCGGTATCTTAATATCAAAACAGGAG 876
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 421 TGGTACGTTCTGCTACTTTAAACCAAGATGGCCAGTCCGTTATTTCTATACAGATTAC 480
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 877 TGGTACGTTCTGCTACTTTAAACCAAGATGGCCAGTCCGTTATTTCTATACAGATTAC 936
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 481 TCAGGTATCTGGAAGATGATGGAACCGGTGCTGTATCAAAATCATTTCACTGCTAA 540
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 937 TCAGGTATCTGGAAGATGATGGAACCGGTGCTGTATCAAAATCATTTCACTGCTAA 996
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 541 GTAAACTTATCCAGCGGATGACAGTCACTTAAGTGAAGTGCATGATCATATAA 600
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 997 GTAAACTTATCCAGCGGATGACAGTCACTTAAGTGAAGTGCATGATCATATAA 1056
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 601 TCTGCTTTGATGGCGGAGACGGTACATTATCAAAATATTCAGCAATTTATCGATGAA 660
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1057 TCTGCTTTGATGGCGGAGACGGTACATTATCAAAATATTCAGCAATTTATCGATGAA 1116
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 661 GGCAAGTGGATTCAGGTATACCATCTTTAAGAGCCCTCACTATGTTGAAGATTAAG 720
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1117 GGCAAGTGGATTCAGGTATACCATCTTTAAGAGCCCTCACTATGTTGAAGATTAAG 1176
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 721 GGCCATAAATATCTTGTCTTTGAAGCGAATCTGGAACAACAGATGTTATCAAGCGCAT 780
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1177 GGCCATAAATATCTTGTCTTTGAAGCGAATCTGGAACAACAGATGTTATCAAGCGCAT 1236
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 781 CAGTCTTCAATTAATAAGCTTACTATGCGGAAGTGAAGTCTTCTTCAGAAATGAAAA 840
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1237 CAGTCTTCAATTAATAAGCTTACTATGCGGAAGTGAAGTCTTCTTCAGAAATGAAAA 1296
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 841 AATAAATCGTTCAAAAGTCTTAATAAACAATTTGCTTTTGAAGATGTTGATGAGC 900
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1297 AATAAATCGTTCAAAAGTCTTAATAAACAATTTGCTTTTGAAGATGTTGATGAGC 1356
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 901 ATTGTTGAATGGCGGATGACTATACAGTGAAGTGTATGAACCATTTAGTGCATCA 960
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1357 ATTGTTGAATGGCGGATGACTATACAGTGAAGTGTATGAACCATTTAGTGCATCA 1416
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 961 AACACAGTAGCAGATGAAGTGAAGCGCCCAATATATTAAATCAATTAATTAATGAT 1020
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1417 AACACAGTAGCAGATGAAGTGAAGCGCCCAATATATTAAATGAATTAATTAATGAT 1476
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1021 CTATTCACGATTCAGAGGATCCAAAATGACAGTGAATTAACGACAAAGATGTT 1080
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 1477 CTATTCACGATTCAGAGGATCCAAAATGACAGTGAATTAACGACAAAGATGTT 1536
OY 1081 TATATGCTAGGCGCGGAGCGGACTCCTTAATGAGCCACACACCGGATTAATGAAGT 1140
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1537 TATATGCTAGGCGCGGAGCGGACTCCTTAATGAGCCACACACCGGATTAATGAAGT 1596
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1141 GGACTTGTATGACATGAATCTTGACCTGCTGATCTCAGACACACTTACTCATTTGC 1200
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1597 GGACTTGTATGACATGAATCTTGACCTGCTGATCTCAGACACACTTACTCATTTGC 1656
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1201 GGTATCCCGCACCCCTGAAGTAAATATGTTGTTACTCACAAGTATATGACGAATGAGCC 1260
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1657 GGTATCCCGCACCCCTGAAGTAAATATGTTGTTACTCACAAGTATATGACGAATGAGCC 1716
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1261 TTCTATCCAGAACATCACTCTCACCTGCGGAGCAAGCTTGCGTTAATTAAGGCT 1320
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1717 TTCTATCCAGAACATCACTCTCACCTGCGGAGCAAGCTTGCGTTAATTAAGGCT 1776
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1321 GACACATCTGAGAGAAAATAGTTCCGAGACAGACAAATTCACA 1365
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1777 GACACATCTGAGAGAAAATAGTTCCGAGACAGACAAATTCACA 1821
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-09-317-179-5
: Sequence 5, Application US/09317179
: Patent No. 6383769
: GENERAL INFORMATION:
: APPLICANT: Tsusaki et al.
: TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/317, 179
: FILING DATE: 24-May-1999
: CLASSIFICATION: <unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/870, 827
: FILING DATE: <unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: TSUSAKI-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2408 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: ORIGINAL SOURCE:
: ORGANISM: Bacillus sp.
: INDIVIDUAL ISOLATE: V230 (FERM BP-5054)
: FEATURE:
: NAME/KEY: signal peptide
: LOCATION: 1..360
: IDENTIFICATION METHOD: E
: SEQUENCE DESCRIPTION: SEQ ID NO: 5:
```

US-09-317-179-5

Query Match 100.0%; Score 1365; DB 4; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACAGCGGGAGCTACAGAGAGCTATGTTTTGCCCATATTTACACGCGCTGACATG 60  
DB 457 ATGACAGCGGGAGCTACAGAGAGAGCTATGTTTTGCCCATATTTACACGCGCTGACATG 516  
QY 61 CTAATAATTCAGAGACACAAACAGCTCTCAATTTAAAGGCCCAATTCATGATCA 120  
DB 517 CTAATAATTCAGAGACACAAACAGCTCTCAATTTAAAGGCCCAATTCATGATCA 576  
QY 121 GCAATCAAAAACATTTGATTCGCAAAAGGATGATAGTACAGCACTTAATAGATT 180  
DB 577 GCAATCAAAAACATTTGATTCGCAAAAGGATGATAGTACAGCACTTAATAGATT 636  
QY 181 GATGATAGGATAGTGGCCACTGCAAAAACGCTGATGCTGCGCAAAATATGATGA 240  
DB 637 GATGATAGGATAGTGGCCACTGCAAAAACGCTGATGCTGCGCAAAATATGATGA 696  
QY 241 TATCATGCTGCTCCGCTTTAGCAGGTGACCCAAAACAGATGATGATCTCAT 300  
DB 697 TATCATGCTGCTCCGCTTTAGCAGGTGACCCAAAACAGATGATGATCTCAT 756  
QY 301 TTATCTATCAAAAAGTGGTATGATGATGATGATGATGATGATGATGATGATGAT 360  
DB 757 TTATCTATCAAAAAGTGGTATGATGATGATGATGATGATGATGATGATGATGAT 816  
QY 361 TTTGAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
DB 817 TTTGAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876  
QY 421 TGGTAGGTTGCTGCTACTTAACCAAGATGCGCAAGTCCGTTTCTATACATATAC 480  
DB 877 TGGTAGGTTGCTGCTACTTAACCAAGATGCGCAAGTCCGTTTCTATACATATAC 936  
QY 481 TCAGTAAATCTGAAAGATGATGAAACCGTCTGTAACCAAAATATTCACACTGCTAA 540  
DB 937 TCAGTAAATCTGAAAGATGATGAAACCGTCTGTAACCAAAATATTCACACTGCTAA 996  
QY 541 GTAACCTATCCACCGCGATGACCTAATAAGTCGATGATGATGATGATGATGAT 600  
DB 997 GTAACCTATCCACCGCGATGACCTAATAAGTCGATGATGATGATGATGATGAT 1056  
QY 601 TCTGCTTTGATGGGAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 660  
DB 1057 TCTGCTTTGATGGGAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 1116  
QY 661 GGCAAGTGGATTTGAGTATACCATCTTTAAGAGACCCCTACATATGTAAGATAG 720  
DB 1117 GGCAAGTGGATTTGAGTATACCATCTTTAAGAGACCCCTACATATGTAAGATAG 1176  
QY 721 GGCCATTAATATCTTGTCTTTGAAGCGAATCTGACACACAGATGTTATCAGGCAT 780  
DB 1177 GGCCATTAATATCTTGTCTTTGAAGCGAATCTGACACACAGATGTTATCAGGCAT 1236  
QY 781 CAGTCTTCATTAATTAAGCTTACTATGCGGAAGTACGCTCTCTCCAGATGAATAA 840  
DB 1237 CAGTCTTCATTAATTAAGCTTACTATGCGGAAGTACGCTCTCTCCAGATGAATAA 1296  
QY 841 AATAAACTGCTTCAAAAGTCTTAATAAACCAATTTCTTTAGCAATGATGATGATG 900  
DB 1297 AATAAACTGCTTCAAAAGTCTTAATAAACCAATTTCTTTAGCAATGATGATGATG 1356  
QY 901 ATTGTTGAATGGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
DB 1357 ATTGTTGAATGGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1416  
QY 961 AACACAGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
DB 1417 AACACAGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1476

QY 1021 CTATTACAGGATTTCAAGAGATATCCAAATGACAGATGATGATGATGATGATGAT 1080  
DB 1477 CTATTACAGGATTTCAAGAGATATCCAAATGACAGATGATGATGATGATGATGAT 1536  
QY 1081 TATATGTAAGGCGCCGAGAGGACTCTTAATATGAGCCACAGAACCCGATTAATGAA 1140  
DB 1537 TATATGTAAGGCGCCGAGAGGACTCTTAATATGAGCCACAGAACCCGATTAATGAA 1596  
QY 1141 GGACTTGTATTAACATATCTTTGACCTGCTGATCTCACACACACTTACTTCATTGC 1200  
DB 1597 GGACTTGTATTAACATATCTTTGACCTGCTGATCTCACACACACTTACTTCATTGC 1656  
QY 1201 GGTATCCCGCACCCGAGAGTAAATATGATGATGATGATGATGATGATGATGATG 1260  
DB 1657 GGTATCCCGCACCCGAGAGTAAATATGATGATGATGATGATGATGATGATGATG 1716  
QY 1261 TTCTATCCAGAAACATCACTCTCACCTCGCGGACAGCTTGGGTTAATATTAAGGCTCT 1320  
DB 1717 TTCTATCCAGAAACATCACTCTCACCTCGCGGACAGCTTGGGTTAATATTAAGGCTCT 1776  
QY 1321 GACACATCTGGAGGAGAAATAGTTCCGACACAGACATTCCTCA 1365  
DB 1777 GACACATCTGGAGGAGAAATAGTTCCGACACAGACATTCCTCA 1821

RESULT 5  
US-09-058-746-1/C  
Sequence 1, Application US/09058746  
Patent No. 6022716  
GENERAL INFORMATION:  
APPLICANT: Ilya Chumakov  
APPLICANT: Hiroaki Tanaka  
TITLE OF INVENTION: High Throughput DNA sequencing vector  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 501 West Broadway  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-3505  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Win95  
SOFTWARE: Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/058, 746  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: GENSET, 015A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10317 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: DOUBLE  
TOPOLOGY: CIRCULAR  
MOLECULE TYPE: synthetic DNA  
ORIGINAL SOURCE:  
ORGANISM: Cloning vector pGendel  
FEATURE:  
NAME/KEY: pGendel  
LOCATION: 1..10317  
FEATURE:  
NAME/KEY: Homology with X06404 comp1 (411..1668)  
LOCATION: 9..1266

IDENTIFICATION METHOD: blastn against X06404  
FEATURE:  
NAME/KEY: Kanamycin resistance gene CDS  
LOCATION: 142..957  
IDENTIFICATION METHOD: By homology to X06404  
FEATURE:  
NAME/KEY: Tn1000's right end  
LOCATION: complement 1332..1371  
IDENTIFICATION METHOD: blastn against X06200)  
FEATURE:  
NAME/KEY: Homology with U46017 (1-472)  
LOCATION: 1423..1894  
IDENTIFICATION METHOD: blastn against U46017  
FEATURE:  
NAME/KEY: single stranded DNA replication origin  
LOCATION: 1423..1894  
IDENTIFICATION METHOD: By homology to U46017  
OTHER INFORMATION: mutation T -> C 1658  
FEATURE:  
NAME/KEY: Homology with U5113 (2382..6997)  
LOCATION: 1896..6544  
IDENTIFICATION METHOD: blastn against U5113  
FEATURE:  
NAME/KEY: OriS  
LOCATION: 1972..2188  
IDENTIFICATION METHOD: By homology to U5113  
FEATURE:  
NAME/KEY: repELR  
LOCATION: 2897..2918  
OTHER INFORMATION: Described in seqid 16  
FEATURE:  
NAME/KEY: Repe  
LOCATION: 2903..3034  
IDENTIFICATION METHOD: By homology to U5113  
FEATURE:  
NAME/KEY: T3  
LOCATION: 3043..3059  
OTHER INFORMATION: Described in seqid 17  
FEATURE:  
NAME/KEY: LRT3RA  
LOCATION: complement 3045..3069  
OTHER INFORMATION: Described in seqid 15  
FEATURE:  
NAME/KEY: Incc  
LOCATION: 3070..3320  
IDENTIFICATION METHOD: By homology to U5113  
OTHER INFORMATION: insertion 33 bases 3038..3071  
FEATURE:  
NAME/KEY: Para  
LOCATION: 3655..4821  
IDENTIFICATION METHOD: By homology to U5113  
OTHER INFORMATION: mutation G -> A 3878  
FEATURE:  
NAME/KEY: ParaB  
LOCATION: 4821..5792  
IDENTIFICATION METHOD: By homology to U5113  
FEATURE:  
NAME/KEY: ParaC  
LOCATION: 5865..6382  
IDENTIFICATION METHOD: By homology to U5113  
FEATURE:  
NAME/KEY: Homology with J01688 (complement 175..819)  
LOCATION: 6574..7218  
IDENTIFICATION METHOD: blastn against J01688  
OTHER INFORMATION: mutation A -> G 7096  
FEATURE:  
NAME/KEY: CDS streptomycin sensitivity gene  
LOCATION: complement 6716..7090  
IDENTIFICATION METHOD: By homology to J01688  
OTHER INFORMATION: mutation A -> G 6728  
OTHER INFORMATION: mutation G -> C 6821  
OTHER INFORMATION: mutation C -> T 6866  
OTHER INFORMATION: mutation T -> C 7013

OTHER INFORMATION: mutation T -> A 7058  
FEATURE:  
NAME/KEY: rpsLR  
LOCATION: 7155..7174  
OTHER INFORMATION: Described in seqid 12  
FEATURE:  
NAME/KEY: SP6  
LOCATION: 7230..7248  
OTHER INFORMATION: Described in seqid 13  
FEATURE:  
NAME/KEY: Tn1000's left end  
LOCATION: 7252..7291  
IDENTIFICATION METHOD: blast (X06200)  
FEATURE:  
NAME/KEY: Homology with X02730 (complement 37..1959)  
LOCATION: 7305..9227  
IDENTIFICATION METHOD: blastn against X02730  
FEATURE:  
NAME/KEY: CDS levansucrase gene  
LOCATION: complement 7379..8800  
IDENTIFICATION METHOD: By homology to X02730  
OTHER INFORMATION: mutation T -> C 7466  
OTHER INFORMATION: mutation A -> G 7739  
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347  
OTHER INFORMATION: mutation T -> C 8600  
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772  
FEATURE:  
NAME/KEY: SLR3  
LOCATION: 8711..8731  
OTHER INFORMATION: Described in seqid 14  
FEATURE:  
NAME/KEY: Homology with J01636 (complement 1158..1465)  
LOCATION: 9298..9623  
IDENTIFICATION METHOD: blastn against J01636  
FEATURE:  
NAME/KEY: CDS alpha peptide beta-galactosidase  
LOCATION: complement 9276..9497  
IDENTIFICATION METHOD: By homology to J01636  
FEATURE:  
NAME/KEY: primer HE1  
LOCATION: complement 9465..9479  
FEATURE:  
NAME/KEY: primer HE2  
LOCATION: 9461..9475  
FEATURE:  
NAME/KEY: primer lacRS2AVT  
LOCATION: complement 9603..9630  
FEATURE:  
NAME/KEY: primer lacE2Mlu  
LOCATION: 9289..9314  
FEATURE:  
NAME/KEY: Homology with M77789 (1889..2576)  
LOCATION: 9629..10315  
IDENTIFICATION METHOD: blastn against M77789  
FEATURE:  
NAME/KEY: high copy-number double-stranded DNA replication origin  
LOCATION: complement 9629..10315  
IDENTIFICATION METHOD: By homology to M77789  
OTHER INFORMATION: mutation C -> T 9803  
OTHER INFORMATION: site ScaI 10029 - 10034  
OTHER INFORMATION: site PmlI 10038 - 10043  
OTHER INFORMATION: CLONING SITES 10031 - 10041  
FEATURE:  
NAME/KEY: oriLRd  
LOCATION: 9856..9881  
OTHER INFORMATION: Described in seqid 8  
FEATURE:  
NAME/KEY: OSI  
LOCATION: 10009..10026  
OTHER INFORMATION: Described in seqid 10  
FEATURE:  
NAME/KEY: ORI  
LOCATION: complement 10046..10062



OTHER INFORMATION: Described in seqid 11  
 FEATURE:  
 NAME/KEY: ORILRT  
 LOCATION: complement 10182..10202  
 OTHER INFORMATION: Described in seqid 9  
 US-09-058-746-1

Query Match 42.3%; Score 577; DB 3; Length 10317;  
 Best Local Similarity 67.2%; Pred. No. 3.5e-152;  
 Matches 902; Conservative 0; Mismatches 390; Indels 51; Gaps 4;

QY 1 ATGACAGCGGGGACTACAAAGAGATATGTTTCCCATATTAACGCGCTGACATG 60  
 DB 8707 ACGAACCAAAAGCCATTAAGGAAACATAGCGCATTTCCCATATTAACGCGCATGATG 8648  
 QY 61 CTAAAAATTCAGAGACAAACACAGTCTCAATTTAAAGGCGCTCATTCATCATCA 120  
 DB 8647 CTGCAATCCCTGACAGCAGCAAAAAATGAATAATATCAAGTTCTTGAGTTGCTCC 8588  
 QY 121 GCAATCAAAAACATTTGATGCGCAAAAGGATATGATAGTCAAGCACTTAATGATTTA 180  
 DB 8587 ACAATTAATAATCTCTCTCGCAAAAG-----CCTG 8555  
 QY 181 GATGATAGGATAGCTGCGCCATGCAAAAGCGTATGCTACTGCGCAATTAATCATGGA 240  
 DB 8554 GAGCTTTGGACAGCTGGCCATTAACAAACGCTGACGCGCATCTGCAAACTATACGCGC 8495  
 QY .241 TATCACATCGTCCGCTTATAGCAGTGAACCAAAAAACAGTATGATCTCCACTCAT 300  
 DB 8494 TACCACATCTCTTTGATTAACCGGAGATCTTAATAATGCGGATGACACATCATTTAC 8435  
 QY 301 TTATTTATCAAAAAGTGGTGATATCATGATGACAGCTGGAATAATGCTGGAAGATA 360  
 DB 8434 ATGTTTATCAAAAAGTGGCGGAAATCTTATGACAGCGGAATAATGCGTGGCGGCTC 8375  
 QY 361 TTGGAAGATATGATTAATTTGTTCCAAATGATCCGATCTTAATATCAACACAGAG 420  
 DB 8374 TTTAAAGACGAGCAAAATTCAGATGATCTCTTAAGAACCAACCAAGCAAGAA 8315  
 QY 421 TGGCAGGTTCTGCTACTTTAACCAAGATGGCCAGTGGTATCTTATACAGATTAC 480  
 DB 8314 TGGCAGGTTCTGACGACATTTACATCTGACGGAATAATCCGTTTATCTACACTGATT 8255  
 QY 481 TCAGATTAATCTGAAGATGTGGAACCGGTGCTGTTAACCAAAATCAATTCACATGCTCA 540  
 DB 8254 TCCGGTAAACATTA-----CGGCAACAAACACATGACACATGACCA 8213  
 QY 541 GTAAACTTATCCAGCGCGATGACGCTACACTTAAAGTCGATGAGATCTGATCATAA 600  
 DB 8212 GTTAACGATACAGCATAGACAGCTCT--TTGAACATCAACGGTGTAAAGGATTATAA 8156  
 QY 601 TCTGCTTTGATGGCGGAGAGGATACAGTTTATCAAAATATTCAGCAATTTATGATGAA 660  
 DB 8155 TCAATCTTTA---CGGTGACGGAACCAACGATATCAAAATGTACAGCGATTCATGATGAA 8099  
 QY 661 GGCAAGTGAATTCAGGTGATTAACCATATCTTTAGAGACCCCTCATGTTGTAAGATAG 720  
 DB 8098 GGCAACTACAGCTGACGCGACACATACGCTGAGAGATCTCTACGTAGAGATTAATA 8039  
 QY 721 GGCCATTAATATCTGCTTTGGAAGCAATACAGTGAACACAGATGTTTCAAGGCGAT 780  
 DB 8038 GGCACAAATAATCTAGATTTTGAAGCAACATGGAATCAAGATGCTCAACAGGCGAA 7979  
 QY 781 CAGTCTTTCAATATAAGCTTACTATGCGGAAGTACGCTCTTCGCAAGATGAATA 840  
 DB 7978 GATCTTTATTTAAACAAGCATATGCAAAAAGCATCATCTCTCCGTCAGAAAGT 7919  
 QY 841 AATTAAGCTTCAAAAGTCTCAAAAACAAATGCTTTTAGGAAGTGTGATGAGC 900  
 DB 7918 CAAAAGCTTCTGCAAAAGCATTAATAAAGCAGCGCTAGTAGAAGCGGCGCTCTCGGT 7859  
 QY 901 ATTGTTAATTTGGCCGATGATACAGTGAATAAGTGTATGAACCATTAATGTCATCA 960

DB 7858 ATGATGAGCTAAACGATGATTACACACTGAAAAAGTGATGAACCGCTGATTCATCT 7799  
 QY 961 AACACAGTACGATGAGTGAAGTGCAGCGCCCAATATTTAAATGANTAAATGAT 1020  
 DB 7798 AACACAGTACGATGAGTGAAGTGCAGCGCCCAATCTTTAAATGAGCGCAATGAT 7739  
 QY 1021 CTATTCACGATTCAGAGATTCGCAAAATGACAGGATGATGAATTAAGCAAAAGATGT 1080  
 DB 7738 CTGTTCACTGACTCCCGCGGATCAAAAATGACATGATGACGCAATGCTGATACGATAT 7679  
 QY 1081 TATATGCTAGGCGCCGAGCGACTCTTAATATGCGCCACACAAACCGATTAATGAAC 1140  
 DB 7678 TACATGCTGATGATTTGTTCTTAATCTTAATCTGAGCCCATACACCGCTGAACAAAC 7619  
 QY 1141 GGACTGATTAATGAACATGATCTGACCGCTGATCTGACACAGCTTACTCTCATTTG 1200  
 DB 7618 GGCCTTGTGTTAAATGATCTTGATCTGATACGATGATTAACCTTACTTACTCACA 7559  
 QY 1201 GGTATCCGCAACCTGGAAGTAAATATGTTGTTACTCAAGATTATAGCAATAGAGC 1260  
 DB 7558 GCTGTACTCTCAAGGAAAGAAACATGCTGATTAACAAGCTATATGACAAACAGAGA 7499  
 QY 1261 TTTATCCAGACATCTCTCACTGCGGACAGAGCTTGGGTTAATTAATGAGCTCT 1320  
 DB 7498 TTTTACGACAGCAAAACATCAACGTTTGCGCGAGCTTCTGCTGAGACATCAAG 7439  
 QY 1321 GACACATCTGGAGGAGAAATAG 1343  
 DB 7438 AAAACATCTGTTGTCAAGACAG 7416

## RESULT 6

US-09-438-142-1/c  
 Sequence 1, Application US/09438142

Patent No. 6258571  
 GENERAL INFORMATION:

APPLICANT: Ilya Chumakov  
 APPLICANT: Hiroaki Tanaka  
 TITLE OF INVENTION: High Throughput DNA sequencing vector  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Knodbe, Martens, Olson & Bear  
 STREET: 501 West Broadway  
 CITY: San Diego  
 STATE: California  
 COUNTRY: USA  
 ZIP: 92101-3505

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: Win95

SOFTWARE: Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/438,142

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Israelson, Ned A.  
 REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: GENSET.015C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-0176  
 TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 10317 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: synthetic DNA  
 ORIGINAL SOURCE:

ORGANISM: Cloning vector pGendel  
 FEATURE: NAME/KEY: pGendel

```
LOCATION: 1..10317
FEATURE:
NAME/KEY: Homology with X06404 compl (411..1668)
LOCATION: 9..1266
IDENTIFICATION METHOD: blastn against X06404
FEATURE:
NAME/KEY: Kanamycin resistance gene CDS
LOCATION: 142..957
IDENTIFICATION METHOD: By homology to X06404
FEATURE:
NAME/KEY: Tn1000's right end
LOCATION: complement 1332..1371
IDENTIFICATION METHOD: blastn against X60200)
FEATURE:
NAME/KEY: Homology with U46017 (1-472)
LOCATION: 1423..1894
IDENTIFICATION METHOD: blastn against U46017
FEATURE:
NAME/KEY: single stranded DNA replication origin
LOCATION: 1423..1894
IDENTIFICATION METHOD: By homology to U46017
OTHER INFORMATION: mutation T -> C 1658
FEATURE:
NAME/KEY: Homology with U51113 (2382..6997)
LOCATION: 1896..6544
IDENTIFICATION METHOD: blastn against U51113
FEATURE:
NAME/KEY: OriS
LOCATION: 1972..2188
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: repELR
LOCATION: 2897..2918
OTHER INFORMATION: Described in seqid 16
FEATURE:
NAME/KEY: RepE
LOCATION: 2903..3034
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: T3
LOCATION: 3043..3059
OTHER INFORMATION: Described in seqid 17
FEATURE:
NAME/KEY: LRT3RA
LOCATION: complement 3045..3069
OTHER INFORMATION: Described in seqid 15
FEATURE:
NAME/KEY: Incc
LOCATION: 3070..3320
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: Insertion 33 bases 3038..3071
FEATURE:
NAME/KEY: Para
LOCATION: 3655..4821
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: mutation G -> A 3878
FEATURE:
NAME/KEY: Para
LOCATION: 4821..5792
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: ParC
LOCATION: 5865..6382
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: Homology with J01688 (complement 175..819)
LOCATION: 6574..7218
IDENTIFICATION METHOD: blastn against J01688
OTHER INFORMATION: mutation A -> G 7096
FEATURE:
NAME/KEY: CDS streptomycin sensitivity gene
LOCATION: complement 6716..7090
IDENTIFICATION METHOD: By homology to J01688

OTHER INFORMATION: mutation A -> G 6728
OTHER INFORMATION: mutation G -> C 6821
OTHER INFORMATION: mutation T -> C 6866
OTHER INFORMATION: mutation T -> C 7013
OTHER INFORMATION: mutation T -> A 7058
FEATURE:
NAME/KEY: rpsLR
LOCATION: 7155..7174
OTHER INFORMATION: Described in seqid 12
FEATURE:
NAME/KEY: SP6
LOCATION: 7230..7248
OTHER INFORMATION: Described in seqid 13
FEATURE:
NAME/KEY: Tn1000's left end
LOCATION: 7252..7291
IDENTIFICATION METHOD: blast (X60200)
FEATURE:
NAME/KEY: Homology with X02730 (complement 37..1959)
LOCATION: 7305..9227
IDENTIFICATION METHOD: blastn against X02730
FEATURE:
NAME/KEY: CDS levansucrase gene
LOCATION: complement 7379..8800
IDENTIFICATION METHOD: By homology to X02730
OTHER INFORMATION: mutation T -> C 7466
OTHER INFORMATION: mutation A -> G 7739
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347
OTHER INFORMATION: mutation T -> C 8600
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772
FEATURE:
NAME/KEY: SLR3
LOCATION: 8711..8731
OTHER INFORMATION: Described in seqid 14
FEATURE:
NAME/KEY: Homology with J01636 (complement 1158..1465)
LOCATION: 9298..9623
IDENTIFICATION METHOD: blastn against J01636
FEATURE:
NAME/KEY: CDS alpha peptide beta-galactosidase
LOCATION: complement 9276..9497
IDENTIFICATION METHOD: By homology to J01636
FEATURE:
NAME/KEY: Primer HE1
LOCATION: complement 9465..9479
FEATURE:
NAME/KEY: primer HE2
LOCATION: 9461..9475
FEATURE:
NAME/KEY: primer LacRS2AVr
LOCATION: complement 9603..9630
FEATURE:
NAME/KEY: primer LacE2Mlu
LOCATION: 9289..9314
FEATURE:
NAME/KEY: Homology with M77789 (1889..2576)
LOCATION: 9629..10315
IDENTIFICATION METHOD: blastn against M77789
FEATURE:
NAME/KEY: high copy-number double-stranded DNA replication origin
LOCATION: complement 9629..10315
IDENTIFICATION METHOD: By homology to M77789
OTHER INFORMATION: mutation C -> T 9803
OTHER INFORMATION: site ScaI 10029 - 10034
OTHER INFORMATION: site PmlI 10038 - 10043
OTHER INFORMATION: CLONING SITES 10031 - 10041
FEATURE:
NAME/KEY: oriLRd
LOCATION: 9856..9881
OTHER INFORMATION: Described in seqid 8
FEATURE:
NAME/KEY: OS1
LOCATION: 10009..10026
```

```

OTHER INFORMATION: Described in seqid 10
FEATURE:
NAME/KEY: OR1
LOCATION: complement 10046..10062
OTHER INFORMATION: Described in seqid 11
FEATURE:
NAME/KEY: orlRr
LOCATION: complement 10182..10202
OTHER INFORMATION: Described in seqid 9
US-09-438-142-1

Query Match      42.38; Score 577; DB 4; Length 10317;
Best Local Similarity 67.28; Pred. No. 3,5e-152;
Matches 902; Conservative 0; Mismatches 390; Indels 51; Gaps 4;

OY 1 ATGAAAGCGGGGAGCTACAGAGAGCTATGCTTTGGCCATTTACAGCGCGGACATG 60
DB 8707 ACGAACCAAAAGCCATATAGGAACATACGGCATTTCCATTTACACGCCATGATATG 8648
OY 61 CTAAAAATCCAGAGACAACAACAGTCTTCATTTTAAAGTGCCTCAATTCATGATCA 120
DB 8647 CTGAAATCCCTGAACGCAAAAAAATGAAATATCAAGTCTCGAGTTGATTCGTC 8588
OY 121 GCAATCAAAACATTTGATTCGGCAAAAGGATATGATAGTACGCAACTAATGATTTA 180
DB 8587 ACAATTTAAATATCTCTTCGCAAAAG-----CCTG 8555
OY 181 GATGATAGGATTTACTGGCCACCTGCAAAAGCTGATGCTAGCGCAATATCATGGA 240
DB 8554 GACCTTTGGAGACCTGGCCCTTACAAACGCTACGCGCATTCGCAACTATACGCG 8495
OY 241 TATACATCGTCTCCGCTTTAGCAGTGAACCAAAACAGATGATCTCACTTCAT 300
DB 8494 TACCACATCGCTTTGATTTAGCCGAGATCTTAAATAGCGAGACATTCATTTAC 8435
OY 301 TTATCTATCAAAAAGTCGATGATACATCATTTACAGCTGGAAAAATGCTGGAAGCTA 360
DB 8434 ATGTTCTATCAAAAAGTCGCGGAACTTCTATTCAGACCTGGAAGAAAGCTGCGCGCTC 8375
OY 361 TTGAAATATGATTAATTTGTCCAAATGATCCGATCTTAAATATCAACAGAG 420
DB 8374 TTTTAAACACGCGCAATTCGATGACAGATGATTCCTTAAAGACCAACCAAGAA 8315
OY 421 TGGTCAAGTCTCTGCTACTTTAACCAGAGTGGCCAGTCCGTTATCTATACAGATTAC 480
DB 8314 TGGTCAAGTCTCAGCCACATTTTACATCTGAGGAAAAATCCGTTATCTACAGATTTC 8255
OY 481 TCAGTATCTCTGAAGATGGTGAACCGGTGCTGTAACCAATCATTTCACTGCTCAA 540
DB 8254 TCCGTTAAACATTA-----CGGCAAAACAAACACTGACACTGACAA 8213
OY 541 GTAAACTTATCCGCGGATGACGCTACACTTAAAGCGATGATTCGATCATTA 600
DB 8212 GTTATACGATACAGATACAGACTCT---TTGACATCAACGGTGTAGAGATTATTA 8156
OY 601 TCTGCTTTTATGAGCGAGAGCGTACAGTTTATCAAAATATTCAGCAATTTATGAGTAA 660
DB 8155 TCATCTTTGA---CGGTGAACGGAACCAATGATCAAAATGATGACAGTTATGATGAA 8099
OY 661 GCGAAGTGGATTTGAGTATACCATCTTTAAGAGACCTCACTATGTTGAAGATTA 720
DB 8098 GCGAATACAGCTCAGCGACCAACCATACGCTGAGAGATCTCTCACTAGAGATTA 8039
OY 721 GGCATTAATATCTTCTTTGAAGGAAATACGTAACCAACAGATGTTATACAGGGAT 780
DB 8038 GGCACAAATACCTTATGTAAGCAAAACACTGAGATGAGTGTACCAAGCGCA 7919
OY 781 CAGTCTTTCAATAATAAGCTTACTATGGGGAAGTACGCTCTCTTCCAGATGAAAA 840
DB 7918 GAATCTTTATTAACAAGACTACTATGCAAAAGCATCATCTTCCGTCAGAAAGT 7919
OY 841 AATTAATGCTTCAAGTCCTAACAAATTCCTTCTTTAGGAAAGGTGCAATTTAGGC 900
DB 841 AATTAATGCTTCAAGTCCTAACAAATTCCTTCTTTAGGAAAGGTGCAATTTAGGC 900

```

```

DB 7918 CAAAACTTTCGCAAGCGATTAACAAAGCCGCTGAGTTAGCAACGGCGCTCTCGCT 7859
OY 901 ATTTGTAATGGCGGATGATCATATACAGTGAAGGTGTTATGAACCATTTAGTCGATCA 960
DB 7858 ATGATTGAGCTTAACGATGATTTACACTGTAAGAAAGTATGAACGCGCTATGCTACT 7799
OY 961 AACACAGTACAGATGAAGTGAAGCGCCCAATATATTTAAATGAATGAATGAATGATAT 1020
DB 7798 AACACAGTACAGATGAATGAAGCGCGCAACGCTTTAAATGAAGCGCAATGATAT 7739
OY 1021 CTATTACAGGATTCAGAGATTCACAAATGACAGTGAATGAATGAATGAATGAATGAT 1080
DB 7738 CTGTTACTGATGCTCCGCGGATCAAAATGACAGATGAGTACGCTCTTACAGATAT 7679
OY 1081 TATATGCTAGGCGCGGAGCGGACCTCTTAAATGAGCCACACCAACCGATTAAGAACT 1140
DB 7678 TACATGCTTGTATGTTCTTAATTTCTTAACGTGCGCATACAGCGCTGACAAACT 7619
OY 1141 GGACTTTATTTGAACATGAATCTTGACCTGCTGATCTACACACACTTACTCTCATTTG 1200
DB 7618 GGCCTGTGTTAAATGATCTTGATCTTACAGATGTAACCTTTACTTACTACACACTTC 7559
OY 1201 GGTATCCGCGACCTGGAAGTAAATATGTTGTTACTACAGATTAATGACAAATGAGGC 1260
DB 7558 GCTGATCTCAAGGGAAGGAACATGCTGATTTACAACTATATGACAAACAGAGA 7499
OY 1261 TTCTATCGAAGAACATCATCTCACCTGCGGAGCAACAGTGGGTTAATTAAGGCTCT 1320
DB 7498 TTCTACGAGCAACAAATCAACGTTTGGCGGACCTTCTGCTGAACTAAGAGCAAG 7439
OY 1321 GACACATCTGGAGGAGAAATAG 1343
DB 7438 AAACATCTGTTGCAAAAGACAG 7416

```

```

RESULT 7
US-09-068-043-1
; Sequence 1, Application US/09068043.
; Patent No. 6048694
;
GENERAL INFORMATION:
APPLICANT: MICHAEL GENE BRAMUCCI
APPLICANT: VASANTHA NGARAJAN
TITLE OF INVENTION: POSITIVE SELECTION
TITLE OF INVENTION: VECTOR FOR BACILLUS SP.
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND
ADDRESS: COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT OFFICE 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/068.043
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/006,201
FILING DATE: NOVEMBER 3, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FELTHAM, S. NEIL
REGISTRATION NUMBER: 36,506
REFERENCE/DOCKET NUMBER: CR-9807
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-6460
TELEFAX: 302-773-0164
;
; INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:  
 LENGTH: 3305 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 US-09-068-043-1

Query Match 39.5%; Score 539.6; DB 3; Length 3305;  
 Best Local Similarity 65.7%; Pred. No. 6,9e-142;  
 Matches 873; Conservative 0; Mismatches 404; Indels 51; Gaps 4;

```

16 TACAAGAGAGTATGTTTTCCTATATACGCGGCTGACATCTAAATTCACAGA 75
1913 TACAAGAGAGTATGTTTTCCTATATACGCGGCTGACATCTAAATTCACAGA 1972
76 CACCAAAACAGTCTCATATTTAAAGTCCCTCAATTCATGATCAGCAATCAAAACATT 135
1973 CAGCAGCAAAACGAAAAATACCAAGTGCCTCAATTCGATCAATCAAGATTTAAATATT 2032
136 GATTGGCGCAAAAGGATGATAGTCAAGCACTTAATGATTTAGATGATGGATGAC 195
2033 GAGTCTGCAAAAGACTT-----GATGTGTCGACAGC 2065
196 TGGCCACTGCAAAAGCTGATGATGCTGCGCAAAATTCATGATGATCAGTCTCC 255
2066 TGGCCGCTGCAAAAGCTGATGATGCTGCGCAAAATTCATGATGATCAGTCTCC 2125
256 GCTTTAGCAGGTGACCCCAAAACAGTATGATGATCTCCATCTTATTTATCAAAA 315
2126 GCTTTAGCAGGTGACCCCAAAACAGTATGATGATCTCCATCTTATTTATCAAAA 2185
316 GTCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375
2186 GTCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2245
376 AAATTTGTTCCAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 435
2246 AAGTTGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2305
436 ACTTTACCAAAAGTGGCCAGTCCGTTTATTTCTATACAGATTACTCAGATTAATCTGAA 495
2306 ACTTTACCAAAAGTGGCCAGTCCGTTTATTTCTATACAGATTACTCAGATTAATCTGAA 2364
496 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 555
2365 -----CGGCAAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2407
556 CCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 615
2408 TCTGAT-----GACACACTCAAAATCAACGAGTGAAGATCACAAAAAGTTTTTGA---C 2461
616 GGAGAGCGTACAGTTATCAAAATATTCAGCAATTTATGATGAGGAGGAGGATTTCA 675
2462 GGAGAGCGTACAGTTATCAAAATATTCAGCAATTTATGATGAGGAGGAGGATTTCA 2521
676 GGTGATTAACCATCTTTAAGAGACCTCTATGTTGAAGATTAAGGAGGAGGATTAATCT 735
2522 GCGGACAAACCATCTCTGAGAGACCTCTCTACGTTGAGAGAGGAGGATTAATCT 2581
736 GTCCTTGAAGCAATTAAGAGACCTCTCTATGTTGAAGATTAAGGAGGAGGATTAATCT 795
2582 GTATTCGAAGCAATTAAGAGACCTCTCTACGTTGAGAGAGGAGGATTAATCT 2641
796 AAAGCTTACTATGCGGAGAGTACGCTCTTCTTCCAGATGAAGAAAAATTAATCTGCTCA 855
2642 AAAGCTTACTATGCGGAGAGTACGCTCTTCTTCCAGATGAAGAAAAATTAATCTGCTCA 2701
856 ACTCCATAAAACAATGCTCTTCTTGAAGATGAGGATTTGAGGATTTGAGGATTTGAGG 915
2702 AGCGTAAAAACAACGATGCTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2761

```

```

QY 916 GATGACTATACAGTGAAGAGTGTATGAACCATTAAGTCGATCAACACAGTAGCAGAT 975
DB 2762 AATGATTTACATTTAAGAGTATGAAGCGGCTGATCAACACAGTACAGAT 2821
QY 976 GAAGTCGAGCGGCAATATATTTAAATGAATTAATGATGATGATGATGATGATGAT 1035
DB 2822 GAATTCGAGCGGCAATATGTTTCAAAATGAAGCGCAATATGATGATGATGATGATGAT 2881
QY 1036 AGAGATTCCAAAATGAGAGTATGAAGCAATTAAGCAAGATGATGATGATGATGATGAT 1095
DB 2882 CGCGTTTAAATGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2941
QY 1096 GGAGGCGACTCTTAAATGAGCGCACACACACACACACACACACACACACACACACACAC 1155
DB 2942 GATCAAACTCTTAAACCGGCTTAAACAGCGGCTGAAGCGGCTGAAGCGGCTGCTGCA 3001
QY 1156 ATGAATCTTGAACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1215
DB 3002 ATGGTCTTGAATCCAAACGATGATGATGATGATGATGATGATGATGATGATGATGAT 3061
QY 1216 GAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1275
DB 3062 AAAGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3121
QY 1276 CACTCTACCTGCGGAGCAAGCTTGGGTTAATATTAAGGAGTGAACATCTGAGAGA 1335
DB 3122 AAGCAACATTTGGCGCAAGCTTCTTAATCAACATCAAAAGCAATAAACATCGGTTGC 3181
QY 1336 GAAATAG 1343
DB 3182 AAAACAG 3189

```

RESULT 8  
 US-08-232-463-14/C  
 ; Sequence 14, Application US/08232463  
 ; Patent No. 5670367  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DORNER, F.  
 ; APPLICANT: SCHEFFELINGER, F.  
 ; APPLICANT: FALKNER, F. G.  
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 ; NUMBER OF SEQUENCES: 52  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 1800 Diagonal Road, Suite 500  
 ; CITY: Alexandria  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22313-0299  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/232.463  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/935.313  
 ; FILING DATE:  
 ; APPLICATION NUMBER: EP 91 114 300.6  
 ; FILING DATE: 26-AUG-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)836-9300  
 ; TELEFAX: (703)836-4109  
 ; TELEX: 899149  
 ; INFORMATION FOR SEQ ID NO: 14:



APPLICATION NUMBER: US 07/965,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/329,154  
 FILING DATE: 25-OCT-1994  
 PRIOR APPLICATION DATA:

FLING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FLING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FLING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPND-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7101 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..7098  
US-08-915-136-9

Query Match 2.7%; Score 36.6; DB 4; Length 7101;  
Best Local Similarity 60.6%; Pred. No. 1.5;  
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 166 AACTTAATAGATTAGATGATGGATAGCTGGCCACGCAAAAGCTGATGCTACGCG 225  
DB 3478 AATTCATAGATTAGTGAATGGAATCTGGAAGATGGAAGGCTTCAGGTCATCTACT 3537  
QY 226 GCAATTAATCATGATATCATCATCTCTCCGCTTTAGCA 264  
DB 3538 GTAACGTATGATATGATACCTCTTTTCAGCACCATCA 3576

RESULT 13  
US-08-957-310-9  
Sequence 9, Application US/08957310  
Patent No. 6365158  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
APPLICANT: Kink, John A.  
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/957,310  
FLING DATE: 23-OCT-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FLING DATE: 24-OCT-1994  
APPLICATION NUMBER: US 08/161,907  
FLING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321

FLING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FLING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPND-01121  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7101 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..7098  
US-08-957-310-9

Query Match 2.7%; Score 36.6; DB 4; Length 7101;  
Best Local Similarity 60.6%; Pred. No. 1.5;  
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 166 AACTTAATAGATTAGATGATGGATAGCTGGCCACGCAAAAGCTGATGCTACGCG 225  
DB 3478 AATTCATAGATTAGTGAATGGAATCTGGAAGATGGAAGGCTTCAGGTCATCTACT 3537  
QY 226 GCAATTAATCATGATATCATCATCTCTCCGCTTTAGCA 264  
DB 3538 GTAACGTATGATATGATACCTCTTTTCAGCACCATCA 3576

RESULT 14  
5231168-1  
Patent No. 5231168  
APPLICANT: DIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;  
VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.  
TITLE OF INVENTION: MALARIA ANTIGEN  
NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/409,658  
FLING DATE: 18-SEP-1989  
SEQ ID NO: 1  
LENGTH: 3095  
5231168-1

Query Match 2.7%; Score 36.4; DB 6; Length 3095;  
Best Local Similarity 42.0%; Pred. No. 1.2;  
Matches 211; Conservative 0; Mismatches 291; Indels 0; Gaps 0;  
QY 477 TTACTCAGTAATCTGAGATGATGGAAACCGGATGATTAAGCAATCTTCACTGC 536  
DB 120 TAAATTAATGATATGATGATGAGGATGTTTGAAGAAAGCTCATATGAAATTTTATC 179  
QY 537 TCAAGTAATCTTATCCAGCCGATGAGTACACTTAAATGATGAGAGTATCTGATCA 596  
DB 180 TGAAGTAAGTAACTGATTAATTAATGAATGAATTTGTTTAATCTGACAAAGTGAAC 239  
QY 597 TAAATCTCTTGTATGCGGAGACGATGATTAATCAAAATATTCAGCAATTTATCA 656  
DB 240 TGAACCTGCTGATGATGAAGATGTTATCTGAAGAAAGCAACCTGAAACCTGAAAA 299  
QY 657 TGAAGCAAGTGAATTCAGTGAATTAACCACTTTAAAGAGACCTGATGTTGAGA 716  
DB 300 TGAAGAAAGTATGATGATGAGAGCTCATGAGAAAGATTTTACTGACAAATGATGA 359  
QY 717 TAAAGGCCATTAATATCTTGTCTTGAAGCAATCTGAAACAAGATGTTATCAAG 776  
DB 360 AGAATCAGGTGAAGTGAATGATGTTGAATGAAGAGGTGATTTGAAAGCACTAATCA 419

OY 777 CGATCAGCTCTTCAATAATAAGCTTACTATGGCGGAAGTACGCTCTTCTTCCAGAAATGA 836  
DB 420 TGAAGAAATTTGACCTGATCAAAATGACTGTGAATTAAGTAAATGAAATTAAGTGAATC 479  
OY 837 AAAAAATAACTGCTTCAAAATCCCTAAAAAACAATTCCTTCTTACGGAATGCTGCAAT 896  
DB 480 AGAAAAAGTGTATCTGAACACGAGCTGAAACATGTGAATAATTTGTATCGAATAAAAGTGTATC 539  
OY 897 AGGCAATGTGAATGGCCGATGACTATACAGTGAATAAGTGTATGAACCATTAAGTGC 956  
DB 540 TGAACCACTGACACGATGAATAATGTATCTGAATAAAAGTACATCCGAACGCTGAACA 599  
OY 957 ATCAACACAGTAGCATGAA 978  
DB 600 TGTAGAAAGTGTATCTGAACAA 621

## RESULT 15.

US-09-648-520E-48  
Sequence 48, Application US/09648520E  
Patent No. 6432649

## GENERAL INFORMATION:

APPLICANT: Stich, Roger W.  
APPLICANT: Rikihisa, Yasuko  
TITLE OF INVENTION: Methods for Detecting Ehrlichia Canis and Ehrlichia Chaffeensis  
TITLE OF INVENTION: Vertebrate and Invertebrate Hosts  
FILE REFERENCE: 22727/04069  
CURRENT APPLICATION NUMBER: US/09/648,520E  
CURRENT FILING DATE: 2000-08-25  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 48  
LENGTH: 849  
TYPE: DNA

ORGANISM: ORF sequence, Ehrlichia chaffeensis p28  
US-09-648-520E-48

Query Match 2.6%; Score 35.4; DB 4; Length 849;  
Best Local Similarity 33.6%; Pred. No. 1.4;

Matches 196; Conservative 61; Mismatches 326; Indels 0; Gaps 0;

OY 204 GCAAAACCGTGTACTCGGGCAAAATATCATGATACATGCTCCGCTTTAGC 263  
DB 204 GCAARATTGGGAGGAGGATAYCYMACWCYHMYMSRAHMTVYATTAAGTGTATC 263  
OY 264 AGGTACCCAAAAAACAAGTATGATACCTCCCTTATTTCTATCAAAAAGTGTGA 323  
DB 264 AAATATATCTTTTAAATATGAAAAATATATCTTTTATAGGTTTGCAGAGCTATTGGYTA 323  
OY 324 TACATGATTTGACAGCTGAAAAATCTGGAAGATTTTGAAGATATGATAAATTTGT 383  
DB 324 CTCATGATGATGGGCCAAGAATAGAGCTTGAATATCTTATGACATTTGATGTAAAAA 383  
OY 384 TCCAAATATGCTGTATCTTAAATATCAAAACAGAGTGTGAGTCTGCTACTTTAAC 443  
DB 384 TCAAGGTAACTATTAAGAAAGAACCCATAGTATGCTGTATATCTCTMASRSYWC 443  
OY 444 CAAAGATGGCAAGTCGTTTATTTATATAGATATACAGTATCCGTAAGATGTGG 503  
DB 444 ABBACARCATGMSKAGTCARRATATMTTGTCTTAAAAAAAGAAAGGRTTACT 503  
OY 504 AACCGTCTGTGTAACCAATATTTCAACTGCTCAAGTAAACTATCCAGCCGATGC 563  
DB 504 TGACRTATCTTTATGCTGAAGCGCATGCTATGATATATATGATGAAGATACCTTTTC 563  
OY 564 AGCTACACTTAAAGTCGATGAGTATCTGATCATTAATCTGCTTTGATGGCGGAGACGG 623  
DB 564 TCCTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 623  
OY 624 TACAGTTATCAAAATATTTACAGCAATTTATGATGATGATGATGATGATGATGATGAT 683  
DB 624 TAAATTTCTTACCAAGAAAGTATGTTAAGTACTCTATTAAGCCAGAAACCTTCTGT 683

OY 684 CCATACCTTTAGAGAGACCCCTCACTATGTTGAAGATTAAGGCCATTAATATCTTGTGA 743  
DB 684 STTTRTYGCGRCATVTTTCATTAAGTATRGRAGCAATTTAGATATATCTCTCTCT 743  
OY 744 AGCGATATCTGGAACACAGATGTTATCAAGGCGATCAGTCT 786  
DB 744 AATACCTAATGATCAASWCTTTCAGAGAAAGGAAATTAACCT 786

Search completed: June 18, 2003, 08:59:54  
Job time : 66.3117 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 02:50:32 ; Search time 3489.14 Seconds  
(without alignments)  
11177.167 Million cell updates/sec

Title: US-09-986-682b-5

Perfect score: 2408  
Sequence: 1 CCGGGAATAACTAGATTCC.....TACCAATAAATAATGTCATC 2408

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pin:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359.6	14.9	755	17	B49063 RPC11-4121
2	309.6	12.9	623	17	AQ284271
3	307.4	12.8	654	17	AG158197
4	296.8	12.3	687	17	AG145469
5	292.6	12.2	689	17	AG174787
6	290.8	12.1	653	17	AQ311782

7	286.6	11.9	688	17	AG165661	AG165661 Pan trogl
8	284.4	11.8	656	17	AG158505	AG158505 Pan trogl
c 9	283.4	11.8	678	17	AG019951	AG019951 Homo sapi
c 10	271	11.3	711	17	AG001039	AG001039 Homo sapi
11	267	11.1	669	17	AG156831	AG156831 Pan trogl
12	264.8	11.0	381	17	AQ791204	AQ791204 HS_4555_B
c 13	258.4	10.7	682	17	AG020006	AG020006 Homo sapi
14	250	10.4	934	9	AL575209	AL575209
15	248.6	10.3	696	17	AG185716	AG185716 Pan trogl
16	233	9.7	683	17	AG167579	AG167579 Pan trogl
17	228.6	9.5	622	17	AG160823	AG160823 Pan trogl
18	226	9.4	508	17	AQ317303	AQ317303 RPC11-10
19	220.8	9.2	1101	17	CNS0085C	AL068761 Drosophila
20	217.4	9.0	687	17	AG145595	AG145595 Pan trogl
21	215.8	9.0	488	17	AQ15317	AQ15317 RPC11-10
22	215.2	8.9	661	17	AG019210	AG019210 Homo sapi
23	215.2	8.9	678	17	AG020005	AG020005 Homo sapi
24	208.2	8.6	694	17	AG145903	AG145903 Pan trogl
25	205.4	8.5	694	17	AG169167	AG169167 Pan trogl
26	203.8	8.5	698	17	AG165439	AG165439 Pan trogl
27	203.6	8.5	682	17	AG165125	AG165125 Pan trogl
28	199.6	8.3	674	17	AG165337	AG165337 Pan trogl
29	199.6	8.3	696	17	AG177482	AG177482 Pan trogl
30	193	8.0	601	17	P859L	AG130538 Leishmani
31	192.8	8.0	698	17	AG180444	AG180444 Pan trogl
32	191.4	7.9	683	17	AG168295	AG168295 Pan trogl
33	189.4	7.9	705	17	AG147097	AG147097 Pan trogl
34	188.2	7.8	676	17	AG145781	AG145781 Pan trogl
35	186.6	7.7	385	17	B4767	B4767 RPC11-1A4
36	182.6	7.6	714	17	AG014172	AG014172 Homo sapi
37	180.6	7.5	613	17	AG064065	AG064065 927P1-13C
38	176.8	7.3	752	17	AG171011	AG171011 Pan trogl
39	176.2	7.3	704	17	AG168501	AG168501 Pan trogl
40	173.6	7.2	729	17	AG014170	AG014170 Homo sapi
41	172.8	7.2	731	17	AG162359	AG162359 Pan trogl
42	167.6	7.0	690	17	AG165865	AG165865 Pan trogl
c 43	166.8	6.9	722	17	AG014192	AG014192 Homo sapi
44	165.6	6.9	686	17	AG167003	AG167003 Pan trogl
45	163.2	6.8	935	17	CNS0083C	AL052150 Drosophila

#### ALIGNMENTS

RESULT 1  
LOCUS B49063  
DEFINITION RPC11-4121.TV RPC1-11 Homo sapiens genomic clone RPC1-11-4121, DNA  
ACCESSION B49063  
VERSION B49063.1 GI:2601300  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 755)  
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden  
,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter  
,J.C.  
TITLE Use of BAC End Sequences for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@igir.org

Clones are derived from the human BAC library RPC1-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from

Research Genetics (info@resgen.com). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
 Seq primer: T7  
 Class: BAC ends.

```

FEATURES
  source
    Location/Qualifiers
      1..755
        /organism="Homo sapiens"
        /db_xref="GDB:7501364"
        /db_xref="taxon:9606"
        /clone="RPCI-11-4121"
        /clone_11b="RPCI-11"
        /sex="Male"
        /cell_type="Lymphocytes"
        /note="Vector: PBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
        RCI11 Human Male BAC Library"
BASE COUNT      262 a      169 c      160 g      164 t
ORIGIN
Query Match      14.9%; Score 359.6; DB 17; Length 755;
Best Local Similarity 69.0%; Pred. No. 1.6e-82;
Matches 523; Conservative 0; Mismatches 229; Indels 6; Gaps 2;

QY 947 CTGAAGATGGTGAACCGGCTGCTGTACCAATTCATTTCACTGCTCACTAATCTGCTTTG
    1111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db CTGACATATCCGGTAAACATTTACGGCAACAAAGCCTGACACAGCGCAGTAAATGTGT 63

QY 1007 CCCAGCGCGATGAGTACACTTAAGTCGATGAGATCTGATCATTAATCTGCTTTG 1066
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 64 CAAATCTGAT---GACACACTCAAAATTCACGAGAGTGAAGATTCACAAAACATTTTGG 120

QY 1067 ATGGCGAGACGGGTACAGTTTATCAAAATATTGAGCAATTTATCGATGAAGCGCAAGTGA 1126
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 121 A---CGAGACGGGAAACATATCAGAACGTTTCAGCAGTTTATCGATGAAGCGCAATTTA 177

QY 1127 TTTCAGGTGATTAACCACTTTTAAAGACCCCTCAGTATGTTGAAGATAAAGGCCATAAAT 1186
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 178 CATCGCGGCAACCAATACGCTGAGAGACCCCTCAGTATGTTGAAGCAAAAGGCCATAAAT 237

QY 1187 ATCTGTCTTTGAAGCAATACTGTAACAACAGATGTTATCAAGCGCATCACTCTTTCA 1246
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 238 ACCTTGTATTGAGACCAACAGCGGAACAGAAACAGGATACCAAGCGCAAGATCTTTAT 237

QY 1247 ATATATTAAGCTTACTATGCGGGAAGTGACGCTCTTCTCCAGATGAAATTAACATGCG 1306
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 298 TTAAACAAAGGCTACTACGGGCGGCAACGAACTTCTCCGTAAGAAAGCAGAGCTTC 357

QY 1307 TTCAAGTCTTAATAAACAATTCGTTCTTTACCGAATGTCATTAAGCATTTGTTGAAT 1366
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 358 AGCAGAGCGCTAATAAAGCGGATGTGAGTTAGCGGAACGGCGCTCGGTATCATAGAGT 417

QY 1367 TGGCCGATGACTATACAGTGAAGAGTGTATGAACCATTAAGCGATCAACACAGTAG 1426
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 418 TAAATATATGATTACATTTGAAAAAGTATGAAGCGCGCATCTTCAACACAGGTAA 477

QY 1427 CAGATGAAGTGAACGCGCCCAATATATTAAATTAATTAATTAATGATATCTATTCACGG 1486
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 478 CTGATGAATTCAGCGCGCAATGTTTCAAAATGAACGCAATGTTACTTGTTCACAG 537

QY 1487 ATTCAAGAGATCCAAATGAGAGATGATGAATTAACGACAAAGATGTTTATATGCTAG 1546
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 538 ATTCAAGCGGTTTAAAAATGACATGATGTATTAATCAAAAGATTTTATACATCTTG 597

QY 1547 GGGCCGAGGACACTCTTAATGAGCCACACACACCGCATTAATGAAGTGAAGTGTAT 1606
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 598 GTTATATCAAACTCTTTTAACCGGCCCTTACAAAGCGCGTGAACAAAGCGCTTGTC 657

QY 1607 TGAACATGAATCTTGACCTGCTGATCTCACACACACTTACTCTCATTTGCGGTATCCCG 1666
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 658 TGAATATGGTCTTGTATCCAAAGAGATGTGACATTTCACTTCTACTCTCGAGTCCCG 717

QY 1667 ACCCTGAAGTATATGTGTCTACACAAAGTATATG 1704
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 718 AAGCCAAAGCAACATGTGTATATCAACAGCTATAG 755
  
```

```

RESULT 2
AO284271
LOCUS
DEFINITION
  AO284271
  RPC11-91P11.TV RPCI-11 Homo sapiens genomic clone RPCI-11-91P11,
  DNA sequence.
ACCESSION
  AO284271
VERSION
  AO284271.1 GI:3910589
KEYWORDS
  GSS.
SOURCE
  human.
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 623)
REFERENCE
  Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
  Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  Use of human BAC End Sequences for Sequence-Ready Map Building
  Unpublished (1998)
  CONTACT: Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: mdadams@tigr.org
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.bufileo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.bufileo.edu/ordering) or from
  Research Genetics (info@resgen.com). BAC end search page:
  http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
  Seq primer: T7
  Class: BAC ends.

FEATURES
  source
    Location/Qualifiers
      1..623
        /organism="Homo sapiens"
        /db_xref="GDB:7534930"
        /db_xref="taxon:9606"
        /clone="RPCI-11-91P11"
        /clone_11b="RPCI-11"
        /sex="Male"
        /cell_type="Lymphocytes"
        /note="Vector: PBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
        RCI11 Human Male BAC Library"
BASE COUNT      209 a      141 c      133 g      139 t      1 others
ORIGIN
Query Match      12.9%; Score 309.6; DB 17; Length 623;
Best Local Similarity 70.7%; Pred. No. 1.6e-69;
Matches 411; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 1140 CCATCTTTAAGAGCCCTCAGTATGTTGAAGTAAAGGCCATTAATCTTGTCTTTGA 1199
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1 CCATACGCTGAGACAGACCTTCACTTCTGTAAGCAAAAGCCCAATTAATCTTGTATTCGA 60

QY 1200 AGGATATCTGGAACAACAGATGTTATCAAGCGATGAGTCTTCAATTAATAAGCTTA 1259
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 61 AGCCAAACAGCGGACGAAACAGGATACCAAGCGCAAGATCTTTATTAACAAAGCGTA 120

QY 1260 CTATGCGGGAAGTACGCTCTTCTCCAGATGAATAAATTAAGTCTTCAAAAGTCTTA 1319
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 121 CTAGCGGCGGCGGACGACACTTTCTTCGTAAGAAACGAGAGCTTCACAGCGGCTTA 180

QY 1320 AAAACAAATTTGCTTTTGAAGCAATGTCATTAAGCATTTGTTGAATTTGGCCGATGACTA 1379
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 181 AAAACCGATGCTGAGTATGACGACAGCGCGCTCGGTATCATAGATTAATTAATGATTA 240

QY 1380 TACAGTGAAGAGTATTAACCATTAAGCATCAACACAGTACAGATGAAGTGA 1439
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 241 CACATTGAAAAAGTATATAGCGGCTGATCATCTTCAACACAGGTAAGTGAATGCA 300

QY 1440 AGCGCCCAATATATTAATAATTAATTAATGATATCTATTCAGGATTCAGAGATTC 1499
  
```

D <sub>b</sub>	301	GCGGCGCAATGTTTTCCAAATGAAGCGCAATTGGTACTCTGTACACTGATTCACGGGTTCC	360
O <sub>y</sub>	1500	CAAAATGACGCGTGATGGAATTAACGACAAGAATGTTATATGCTAGGCGCCGAGGCCGA	15358
D <sub>b</sub>	361	AAAAATGCGATCAGATGGTATTAACTCAAACGATTTTTCATGGTTGGTTATGATCAAA	420
O <sub>y</sub>	1560	CTCCTTAAATGGCCACACACCAGATTAATGAATACGAGCTTGTATGAACATGAATCT	1619
D <sub>b</sub>	421	CTCTNTAACCGGCCTTACAAAGCCGCTAACAAAACAGGGCTGTGCTGCGCAAATGGGCTCT	480
O <sub>y</sub>	1620	TGACCCGTCGATCTMCACACACTTACTCTCATTTGGGGTATCCCGCACCCGTGAAGATA	1679
D <sub>b</sub>	481	TGATCCAAACAGATGTGACATCTACTTACTCTCACTTGCAGATGCCCGCAGCCAAAGGCTAA	540
O <sub>y</sub>	1680	TAAATGTGCTACTCACAACTTATATGACGAATAGAGGCTTCT	1720
D <sub>b</sub>	541	CAATGTGGTTATCTCAACAGCTTCAATGACCAACAGAGGGCTTCT	581

RESULT 3							.
AG158197							
LOCUS							
DEFINITION	654 bp DNA linear GSS_09-JAN-2002						
Pan troglodytes DNA, clone: RP43-02F15.T7,							
sequence.							
ACCESSION	AG158197						
VERSION	AG158197.1						
KEYWORDS	GI:16687875						
SOURCE	GSS.						
Pan troglodytes male lymphocytes DNA, clone_1lb:RPci_43 Chimpanzee							

ORGANISM      Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE AUTHORS	TITLE	JOURNAL	REFERENCE AUTHORS
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Motoiki, Y., Watanabe, H. and Sakaki, Y.	BAC end sequences of Library RPO1-43	Unpublished	2 (pages 1 to 654) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

**JOURNAL.** Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chinpbes@sc.riken.go.jp, URL: <http://npg.gsc.riken.go.jp/>, Tel.:81-45-503-9111, Fax:81-45-503-9170)

**COMMENT** Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

**PRIMERS**

Sequencing: T7  
LIBRARY

Vector	:	pBACE3.6
R.site 1	:	ECORI
R.site 2	:	ECORI

```

/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-023F15.17"
/sex="male"
/cell_type="lymphocytes"
/clone_ilb="RP43 Chimpanzee Male BAC Library
216 a 151 c 140 g 146 t 1 others
BASE COUNT
ORIGIN

```

Query Match	12.8%	Score 307.4	DB 17	Length 654
Best Local Similarity	70.9%	Pred. No. 6e-69		
Matches 421; Conservative	0;	Mismatches 172;	Indels 1;	Gaps 1;

[illegible]

Db 22 TTCCGGGCACACCATACGTACGTCGAGACGCCCTCACTACNCTGTTGAAGACAAAGCCATTAAT 81

Qy 1187 ATCTTGTCTTTGAAGCGCAATACTGGAACACACAGATGGTTATCAAGGGCATCTTTCA 1246

Db 82 ACCTGTATTTGACAGCCACACACGGGAACAGAAACGGATACCAAGCGGAGAAATCTTTAT 141

Qy 1247 ATATTAAGCTTACTATGGCGGAAGTAGACGCTCTTCTCCAGATGAAGAAAAATTAATACGC 1306

Db 142 TTAAACAAAGCCTACTACGCGCGGACAGAACTCTTCCGTAAAGAAAGCCAGAAAGCTTC 201

Qy 1307 TTCAAGTCCTTAAAAAACAAATTTGCTTCTTATGGAATGGATGGCATTAAGCCATTTGAT 1366

Db 202 AGCAGAGCGCTTAAAAAAAGCGCATGCTAGTTAGCGAAAGCGGCCCTCGGTTCTATAGAT 261

Qy 1367 TGGCGATGACTATACAGTGAAGAGTGTATGAAGACATTAGTGCATCAACACAGTAG 1426

Db 262 TAAATTAATGATTACATTTGAAAAAGTAATGAAGCGCGATGATCATCTTCAACACGGTTAA 321

Qy 1427 CAGATGAAGTGAACGGCGCCCATATATTAAATGATTAATTAATGGTATCTATTACCG 1486

Db 322 CTGATGAATGAGACGGCGCGAATGTTTCAAAATGAACGGCAAAATGCTACTGTTCACTG 381

Qy 1487 ATTCAAGAGATCCAAAATGACAGTAGTGATGAATTAACGACACAAGATGTTATATAGTAG 1546

Db 382 ATTCCGCGGTTCAAAAATGACAGATCGATGGTATTAACTCAACAGGATATTTACATGGCTTG 441

Qy 1547 GGCCCGGAGGAGCTCTTAAATAGGCCACACACCGCATTAATAGAACTGAGATCTGTAT 1606

Db 442 GTTATGATTCAAACGCTTTAAACGGGCCCTTCAAGCGCGCTGAACAAAACAGGCTTTGCG 501

Qy 1607 TGAACATGAATCTTGACCCCTGCTGATCTCACACACACTTACTCTCATTTGGCGATCCCGC 1666

Db 502 TGCAAATGGGCTTGATATCCAAACGATGAGATTCACATTCTCAGCTTGGCAGATGGCGCG 561

Qy 1667 ACCCGAAGTAATTAATGTGTACTGCCAACATTTATAGCAATAGAGGGCTTC 1720

Db 562 AAGCCAAAGGCAACAAATGTGGTTATCAACAACTCATATGACAAACAGAGGGCTTCT 615

RESULT 4	AG145469	687 bp	DNA	Linear	GSS 08-JAN-2002
LOCUS	Pan troglodytes DNA, clone: RP43-006M11.T7, genomic survey				
DEFINITION	sequence.				
ACCESSION	AG145469				
VERSION	AG145469.1				
KEYWORDS	GI:16675147				
SOURCE	GSS.				
	Pan troglodytes male lymphocytes DNA, clone_11b:RPC1-43 Chimpanzee				

ORGANISM	DATE INC.	DIAGN. CLONE	INSTR.	000011117.
Pan troglodytes				
Eukaryota, Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:				
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Pan.				
REFERENCE				
1				

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.	BAC end Sequences of Library RPCI-43	Unpublished	2 (bases 1 to 687)	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.	Direct Submission	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

1-7-22 Suehiko-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: [chimpbes@isc.riken.go.jp](mailto:chimpbes@isc.riken.go.jp), URL: <http://hnp.isc.riken.go.jp/>,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library RPT1-43 This BAC  
end was generated during the R&D process and may have higher chance  
of clone tracking errors.  
PRIMERS

Sequencing: T7  
LIBRARY  
Vector : pBACE3.6



```

Db      483 TAAATATGATTCACATTTGAAAGTAATGAAGCCGTCATCCTCAACACGGTAA 542
QY      1427 CAGATGAGTCGACCGCCCAATATATTAAATGATATATATGATCTTATTCAGG 1486
Db      543 CTGATGAAATCGAGCGCGGAAATGTTTCAAAATGAACGGCAATGTACTTCTACTG 602
QY      1487 ATTCAAGAGATCAAAATGACGAGTATGATTAACGCAAAAGATGTTATATGCTAG 1546
Db      603 ATTACAGCGGTTCAAAAATGACGATCGATGATTTAATCAACAAAGATATTACATGCTTG 662
QY      1547 GCGCCGAGCGCAGCTCTTAATAGCC 1573
Db      663 GTATATGATCAAACTCTTAACCGGCC 689

RESULT 6
AO311782 653 bp DNA linear GSS 04-MAY-1999
LOCUS    RPC11-103A24.TV RPC1-11 Homo sapiens genomic clone RPC1-11-103A24,
DEFINITION
DNA sequence.
ACCESSION AO311782
VERSION    AO311782.1 GI:4043531
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 653)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
            Use of human BAC End Sequences for Sequence-Ready Map Building
            Unpublished (1998)
TITLE      Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbeetli@f.org
            Clones are derived from the human BAC library RPC1-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.bufileo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.bufileo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES
            source          1..653
                        Location/Qualifiers
                        /organism="Homo sapiens"
                        /db_xref="GDB:7539191"
                        /db_xref="taxon:9606"
                        /clone="RPC1-11-103A24"
                        /clone_lib="RPC1-11"
                        /sex="Male"
                        /cell_type="Lymphocytes"
                        /note="Vector: pBac3.6; Site_1: EcoRI; Site_2: EcoRI;
                        RPC11 Human Male BAC Library"
BASE COUNT 232 a 145 c 143 g 133 t
ORIGIN
Query Match 12.1%; Score 290.8; DB 17; Length 653;
Best Local Similarity 68.0%; Pred. No. 1.2e-64;
Matches 460; Conservative 0; Mismatches 192; Indels 24; Gaps 3;
QY 787 ATTGACAGTGGAAAAATGCTGGAAGATATTGAAGATATGATAATTGTTCCAAAT 846
Db 2 ATGACAGCTGGAAAAACGGCGCGCTGTCTTTAAAGACAGCATAGTTGCGAGCGCAAC 61
QY 847 GATCCGATCTTAATATCAACACGAGAGTGTCTGAGTTCTGCTACTTTAACCAAGAT 906
Db 62 GATCCGATCTCTGAAGATCAGACGAAGAATGTCGCTTCCGCAACCTTACATCTGAC 121

```

```

QY 907 GGGCAAGTCGGTTATTTATACAGATTAATCTAGGTAATCCGAGATGGTGAACCGGT 966
Db 122 GGAATAATCCGTTTATTTACTACTGACTATTCGGTAACATTA----- 165
QY 967 GCTGTAAACCAATTCATTTCACTGCTCAAGTAACCTTATCCAGCGCATCGCTACGA 1026
Db 166 --CGGCAACAAAGCCCTGACCAACAGCGCAGTAAATGTGCAAAATCTGAT---GACACA 220
QY 1027 CTTAAAGTCGATGAGTATCTGATCAATTAATCTGCTCTTTGATGGGAGACGGTACAGTT 1086
Db 221 CTCAAAATCAACGAGTGGAGATACAAAACGATTTTGA---CGAGACGGAATAACA 277
QY 1087 TATCAAAATATATACGAATTTATGATGTAAGGACGATTCAGTGTATTAACCTACT 1146
Db 278 TATCAGAACGTTTATGAGATTTATGATGTAAGGCAATTAATCAATCGGCAACCATAC 337
QY 1147 TTAAGAGACCCCTCACTATGTTGAAGTAAGGGCCATTAATATCTGCTTTGAAGCGAAT 1206
Db 338 CTGAGAGACCCCTCACTATGTTGAAGTAAGGGCCATTAATATCTGATTCGAGCGCAAC 397
QY 1207 ACTGGAACACAGATGTTATCAAGCGCATCTTTCAATATAAAGCTTACTATGGC 1266
Db 398 ACGGGAACAGAAAACGATACCAAGCGGAAGATCTTTATTTAACAAGCGTACTACGGC 457
QY 1267 GGAAGTGACGCTCTCTCCAGATGAAAAAATAAATCTGTTCAAGTCTCTAAAAACAA 1326
Db 458 GGGCGACAGCACTTCTCCGTAAGAAGCCAAAGCTTATGACAGCGCTTAAAAACGC 517
QY 1327 ATTGCTTCTTTTACGAGATGTTGATGATGATGATGATGATGATGATGATGATGAT 1386
Db 518 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 577
QY 1387 AAAAGTGTATGAACCATTTAGTCGATCAACACAGTACGATGATGATGATGATGATGAT 1446
Db 578 AAAAAGTATGAAGCGCGCTGATCACTTATACACGATGATGATGATGATGATGATGATGAT 637
QY 1447 AATATATTTAATGA 1462
Db 638 AATGTTTCAAAATGA 653

RESULT 7
AG165661 688 bp DNA linear GSS 09-JAN-2002
LOCUS    Pan troglodytes DNA, clone: RP43-033D03.T7, genomic survey
DEFINITION
sequence.
ACCESSION AG165661
VERSION    AG165661.1 GI:16695339
KEYWORDS   GSS.
SOURCE     Pan troglodytes male lymphocytes DNA, clone_lib:RPC1-43 Chimpanzee
            Male BAC Library clone:RP43-033D03.T7.
            Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
            BAC end sequences of Library RPC1-43
            Unpublished
TITLE      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Sueni-ro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpses@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
            Clones are derived from the chimpanzee BAC library RPC1-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
            PRIMERS
JOURNAL
AUTHORS

```





```

Db      590 GACGGAAAAATCCGTTTATCTACACTGACTATTCGGTAAACAT----- 546
Oy      964 GGTGCTGTGTAACCAATATCTTCAACTGCTCAAGTAATCTATCCAGCCGATGCACCT 1023
Db      545 --TACGGCCAAACAAAGCCTGACACAGCAGGATTAATGTGTCAAAATCTGAT---GAC 491
Oy      1024 ACATTTAAAGTCATGATGATATCTGATCAATATCTGTCTTGATGCGGAGACGGTACA 1083
Db      490 ACACCAAAATCAACGAGTGGAGATCACAACAAAGATTTTGA---CGGAGACGGAAAA 434
Oy      1084 GTTATATCAAAATATTCACCAATTTATATCATGATGAAGCAGATTCAGTGATACCAT 1143
Db      433 ACATATCAAGAGTTCACAGCTTATATCGATGAAGGCAATTTATACATCGGCGACACCAT 374
Oy      1144 ACTTTAAGAGACCCCTCATGATGTAAGATGAAGGCCATTAATATCTTGTCTTGAACGC 1203
Db      373 ACGCTGAGAGACCCCTTACTAGCTTCAACACAAAGCCATTAATATCTTGTATTCGAAGC 314
Oy      1204 AATACTGACACACAGATGTTATCAAGCGCATCAGTCTTTCATATTAAGCTTACTAT 1263
Db      313 AACACGGGACAGAAAAGGATACCAAGCGAAGATTTATTTAACAAAGCGTATAC 254
Oy      1264 GCGGAGAGTACGCTTCTTCAGATGAAAAAATAACTGCTTCAAGTCTTAATAAA 1323
Db      253 GCGGCGGCGACAGACTTCTTCGTAAGAAAGCCAGAACTTCAGACGAGCTTAATAAA 194
Oy      1324 CAATTTGCTCTTATGCGAATGATGTCATTAAGCATTTGTAATGGCCGATGACTATACA 1383
Db      193 CGCGATGCTGATGATGCGAAGCGGCGCTCGTATCATAGATTAATATATGATTAACACA 134
Oy      1384 GTGAAAGTGTATGAAACCACTTATGCGCATCAACACAGTAGACGATGAAGTGAACGC 1443
Db      133 TTGAAAAAGATGAAGACCCGCTGATCACTTCAACACGCTAAGTGAATGATGAGCGC 74
Oy      1444 GCCATATATTTAAATATGAATATTAATGTGA 1475
Db      73 GCGAATGTTTCAAAATATGACGCAATGTGA 42

RESULT 11
AG156831 669 bp DNA linear GSS 09-JAN-2002
LOCUS Pan troglodytes DNA, clone: RP43-021J04.T7, genomic survey
DEFINITION
ACCESSION AG156831.1 GI:16686509
VERSION AG156831.1
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphocytes DNA, clone_lib:RP43-43 Chimpanzee
ORGANISM Homo sapiens
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Toki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library RP43-43
JOURNAL Unpublished
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Toki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbes@gs.c.riken.go.jp, URL:http://nbp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RP43-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY Vector : pBACe3.6

```

```

FEATURES
Source
R.Site 1 : ECORI
R.Site 2 : ECORI.
Location/Qualifiers
1..669
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-021J04.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP43-43 Chimpanzee Male BAC Library"
BASE COUNT 227 a 153 c 143 g 146 t
ORIGIN
Query Match 11.1%; Score 267; DB 17; Length 669;
Best Local Similarity 65.5%; Pred. No. 1.9e-58;
Matches 390; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

Oy      1205 ATACTGAAACACAGATGTTATGACGCGATCAGTCTTTCATATTAATGAAGCTTATG 1264
Db      34 ACACGGGACACGAAAACGAGATCCACAGCGGAGAAATCTTATTTAACAAAGCGTATACG 93
Oy      1265 GCGGAGTGAAGCTCTCTTCAGATGAAAAAATAACTGCTTCAAGTCTTAATAAAC 1324
Db      94 GCGCGGCGACGACTTCTTCGTAAGAAAGCCAGAACTTCAGACGAGCGCTTAATAAAC 153
Oy      1325 AATTTGCTTCTTATGCGCAATGCTGCTATTAAGCATTTGTAATGGCCGATGACTATACAG 1384
Db      154 GCGATGCTGATTTAGCGAAGCGCGCCCTCGTATCATAGATTAATATATGATTAACAT 213
Oy      1385 TGAAGAAGTGTATGAACACATTAGTGCATCAACACAGTAGACAGATGAAGTGAACGC 1444
Db      214 TGAAAAAAGTATGAAGCGCGCTGATCACTTCAACACGATGATGAATGATGAGCGC 273
Oy      1445 CCATATATTTAAATATGAATATTAATGATGATGATGATGATGATGATGATGATGATGAT 1504
Db      274 GCAATGTTTCCAAATGACGCGCAATATGATGATGATGATGATGATGATGATGATGATGAT 333
Oy      1505 TGACAGTGTATGATTAACGACCAAGATGTTTATATGATGATGATGATGATGATGATGAT 1564
Db      334 TGACGATGATGATGATTAATCACTCAACAGATATTTATGATGATGATGATGATGATGAT 393
Oy      1565 TAAATGGCCCAACACCCGATTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1624
Db      394 TAACGGCGCTTAACAGCGCTGAAACAAAGCGCTGACGCAATATGATGATGATGATGAT 453
Oy      1625 CTGCTGATGTCACACACACTTACTCTGATTCGCGTATCCGACCGCTGAAGGATATATG 1684
Db      454 CAACAGATGATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 513
Oy      1685 TGTATGTCACAAAGTATATGAGAGGCTTCTATCCAGAAACATCACTTCACTGCG 1744
Db      514 TGTATATCACAGCTACATGACAAAGAGGCTTCTGAGAGATTAAGGACACATTTG 573
Oy      1745 GCGCAAGCTTGGGGTATATTTAAAGGTCGTGACACATCTGGAGGAGAAATATG 1799
Db      574 GCGCAAGCTTGTATATGATGAACATCAAGCAATTAACATCCGTTGCAAAAAACAG 628

RESULT 12
AO791204 581 bp DNA linear GSS 03-AUG-1999
LOCUS HS_4555_B1_A10.T7A CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-4555 Col-19 Row-B, DNA sequence.
ACCESSION AO791204
VERSION AO791204.1 GI:5698916
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Maitiras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

```





```

QY 1708 AATAGAGGCTTCTATCCAGAACATCTACCTCGGGACAAAGCTGGGTTAATATT 1167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 AAGAGAGGCTTCTTCAGAGGATTAAGGACATTTGGCGCAAGCTTTTAATGAACATC 142
QY 1768 AAGAGGCTGCACATCTGGAGAGAGAAATAG 1799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 AAGGCAATFAAACATCCGTTGTCAAAAACAG 110

RESULT 14
AL575209 934 bp mRNA linear EST 16-FEB-2001
LOCUS AL575209 LTI_NFL006_PL2 Homo sapiens cDNA clone CSOD1062YP02 5
DEFINITION prime, mRNA sequence.
ACCESSION AL575209
VERSION AL575209.1 GI:12936153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Homo sapiens: Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLES Li, W.B., Gruber, C., Jesse, J. and Polyes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
CONTACT: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
    source
        1. 934
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CSOD1062YP02"
            /clone_lib="LTI_NFL006_PL2"
            /tissue_type="placenta"
            /note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA
            was enriched with a NotI-oligo(dT) primer. Five prime end
            is enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed by
            Life Technologies. Contact : Feng Liang Life Technologies,
            a division of Invitrogen 9800 Medical Center Drive
            Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
            Email : fliang@life.com URL :
            http://fulllength.invitrogen.com"
BASE COUNT 298 a 196 c 223 g 209 t 8 others
ORIGIN
Query Match 10.4%; Score 250; DB 9; Length 934;
Best Local Similarity 64.8%; Pred. No. 5.1e-54;
Matches 367; Conservative 2; Mismatches 197; Indels 0; Gaps 0;

QY 1234 GATCAGCTTTCAATATAAGCTTACTATGCGGAGAGTACTTCTTCAGAAATGAA 1293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 GAAGAATCTTATTATTAACAAAGCTACTACGGCGGCGACGAATCTTCCGTAAGA 82
QY 1294 AAAAATAAAGCTGCTCAAGTCTTAAACAAATAGCTCTTATGCGAATGCGATTA 1353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 AGCCAGAGCTTAGACGAGCGCTTAATAAAAGCGAGTCTAGTTAGGAACGGCGCTC 142
QY 1354 GGCATTGTGAATTTGGCCGATGACTATACAGTAAAGTGTATGAACCATTTAGTCGA 1413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 GGTATCATAGAGTTAATATATGATTAACACTTGAATAAAGTATGAAGCCGCTGATC 202
QY 1414 TCACACACAGTACAGATGAGTGAACGGCCCAATATTAATAATGAATTAATAG 1473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 TCACACACAGTACAGTGAATGAGCGCGCAATGTTTCAAAATGAACGGCAATAG 262
QY 1474 TATCTATTCAGGATTCAGAGATTCAGCAAAATGACGAGTATTAACGCAAAAGAT 1533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 TACTTGTCTACATTCACGCGGCTCAAAAATGACGATGATTAATTAACCAAGAT 322

```

```

QY 1534 GTTATATGCTAGAGCCCGGAGCGACTCTTAATATGGCCACAGACCCGATTAATGNA 1593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 ATTATATCTCTTGTTATATATATCAACACTCTTAACCGGCCCTTACAAAAMAAAAA 382
QY 1594 ACTGACTTGTATTTAGACATGAACTTGACCCCTGCTGATTCACACACTTACTCTAT 1653
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 AAGAGGCTTGCTGCTGCAAAATGGTCTTGATCCAAAGATGACATTCCTACTCTCAC 442
QY 1654 TGGGATCCCGCACCCCTGAAGTATATGTTAGTACACAGATTAATGACGAATGA 1713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 TTGACAGTCCCGAAGCCAAAGCAAAATGTTGTTATATCAAGCTTACTGACAAACGA 502
QY 1714 GGCCTTATCCAGAACATCCTCTACCTGCGGACAAAGCTTGGGTTAATTAAGG 1773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 GCTTCTTGAGAGATTAAGAACCAATTTGGCCCAAGCTTCTTAATGAAATCAAGGC 562
QY 1774 TCTGACACATCTGGAGAGAAATAG 1799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 563 AATAAACAATCCGTTGTCAAAAACAG 588

RESULT 15
AG185716 696 bp DNA linear GSS 09-JAN-2002
LOCUS AG185716
DEFINITION Pan troglodytes DNA, clone: RP43-059006.T7, genomic survey
sequence.
ACCESSION AG185716
VERSION AG185716.1 GI:16715396
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
Pan troglodytes
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.
TITLES BAC end sequences of Library RPCI-43
JOURNAL Unpublished
2 (bases 1 to 696)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.
TITLES Direct Submision
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenhiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpses@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY
Vector : pBACE3.6
R.site 1 : EcoRI
R.site 2 : EcoRI.
FEATURES
    source
        1. 696
            /organism="Pan troglodytes"
            /db_xref="taxon:9598"
            /clone="RP43-059006.T7"
            /sex="male"
            /cell_type="lymphocytes"
            /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT 237 a 162 c 148 g 146 t 3 others
ORIGIN
Query Match 10.3%; Score 248.6; DB 17; Length 696;
Best Local Similarity 64.6%; Pred. No. 1.2e-53;
Matches 459; Conservative 0; Mismatches 200; Indels 51; Gaps 4;

QY 473 ACAAGAGAGCTATGCTTTGGCCATATTAACAGCGCTGACATGCTTAATAATTCAGAGAC 532

```

```
Db      38 AAAAAAAAAACCGACGTGCGTCTCTCATATATACAGCCATGATATGCTGCAGATCCCTAAAC 97
QY      533 AACAAACAGTCCTCAATTTAAAGTCCTCAATTCATGCTCAGACATCAAAACATG 592
Db      98 AGCAGCAAAACGAAAAATACCAAGTGCCTCAATTCATGATCAATCAACGATTAATAATATG 157
QY      593 ATTCGGCAAAAGGATATGATGATCAGCCACTTAATAGATTTAGATGATGATGATGCT 652
Db      158 AGTCTGCAAAAGGACTT-----GATGTGTGGACAGCT 190
QY      653 GGGCAGTCAAAACCGCTGATGCTACTGCGCAATTTATCATGATATCAGATGCTCCG 712
Db      191 GGGCCCTGCAAAACCGCTGACAGACATGCTGAATACAAACGCTATCACGTTGTCTTG 250
QY      713 CTTTACAGTCAACCAAAACAGTATGATGATCTCAGTCACTTATTTATTCATCAAAAG 772
Db      251 CTCTTCGGGAGAGCCCGAAAGCGCTGATGACACATCANTCTACATGTTTATCAAAAG 310
QY      773 TCGGTGATACATCGATTGACAGCTGAGAAAAATGCTGAGAGATTTGAAGATATGATA 832
Db      311 TCGGCGACAACTCAATCGACAGCTGAGAAAAAGCGGCGCTGTCTTAAAGACAGGATA 370
QY      833 AATTGTTCCAATGATCCGTATCTTAATATCAAAACAGAGAGTGTGAGTCTGCTA 892
Db      371 AGTTGACGCGCAACGATCCGATCTGAAGATCAGACGCAAGATGTCGGTTCGCAA 430
QY      893 CTTTAAACAAAGATGGCCAGTCCGTTTATTTATATACAGTATCTAGGTAACTCGAAG 952
Db      431 CCTTACATCTGACGAAAAATCCGTTTATTTCTACAGTACTATTCGGTTAAACATTA-- 488
QY      953 ATGTGGAACCGGTGCTGTGTAACCAAAATCATTTCAACTGCTCAGTAACTTATCCAGC 1012
Db      489 -----CGGCAAAACAAAGCCTGACAAACAGCGCAGTAATGTGTCAAAAT 532
QY      1013 CGGATGCACTACACTTAAAGTCGATGAGTATCTGATCATTAATCTGCTTGTGATGCG 1072
Db      533 CTGAT--GACACACTCAAAATCAAGGAGTGAAGATCACAAAACGATTTTGA--CG 586
QY      1073 GAGACGTTACAGTTTATCAAAATATTCAAGATTTATCGATGAGCAAGTGATTTAG 1132
Db      587 GAGACGAAAAACATATCAGAAAGTTCACAGTATTATCGATGAGCAATTTATACATCG 646
QY      1133 GTGATACCATATCTTTAAGAGACCCCTCACTATGTTGAAGATAAGGCCAT 1182
Db      647 GCGACAACCATACGCTGAGAGACCCCTCACTACGTTGAGACATAGGCCAT 696
```

Search completed: June 18, 2003, 08:56:50  
Job time : 3490.14 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 07:25:42 ; Search time 365.061 Seconds  
(without alignments)  
9679.375 Million cell updates/sec

Title: US-09-986-682B-5  
Perfect score: 2408  
Sequence: 1 CGGGGAAATAGTAGATTCC.....TACCAATATAATGGTGATC 2408

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1042519 seqs, 733713590 residues  
Total number of hits satisfying chosen parameters: 2085033

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/PCRUS\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2408	100.0	2408	10	US-09-986-682B-5
2	1365	56.7	1365	10	US-09-986-682B-4
3	627	26.0	1449	10	US-09-974-300-2644
4	608.8	25.3	4938	9	US-10-055-794-1
5	608.8	25.3	6525	9	US-10-055-794-3
6	607.2	25.2	10317	10	US-09-849-866-1
7	580.6	24.1	2253	10	US-09-993-292A-19
8	580.6	24.1	8908	10	US-09-993-292A-18
9	135.6	5.6	368	10	US-09-864-761-21746
10	110.4	4.6	882	10	US-09-974-300-735
11	93	3.9	1146	10	US-09-778-844-136
12	80.2	3.3	376	10	US-09-864-761-5018
13	61.2	2.5	4634	10	US-09-995-587A-10
14	52	2.2	839	9	US-10-198-846-5829
15	52	2.2	2367	10	US-09-995-587A-2
16	52	2.2	2394	10	US-09-995-587A-3
17	52	2.0	2592	10	US-09-995-587A-4
18	47	2.0	931	9	US-10-198-846-5362
19	44	1.8	58985	9	US-09-901-152-3

20	41.8	1.7	5301	7	US-08-781-986A-443	Sequence 443, App
21	39	1.6	394	10	US-09-878-574-3784	Sequence 3784, Ap
22	38.6	1.6	921	10	US-09-815-242-4588	Sequence 4588, Ap
23	38.6	1.6	957	10	US-09-815-242-8509	Sequence 8509, Ap
24	38.4	1.6	5086	9	US-09-764-881-9794	Sequence 9794, Ap
25	38	1.6	314	10	US-09-833-381-190	Sequence 190, App
26	38	1.6	504	9	US-10-074-095-200	Sequence 200, App
27	38	1.6	548	10	US-09-764-860-200	Sequence 200, App
28	38	1.6	548	10	US-09-974-300-3392	Sequence 3392, Ap
29	38	1.6	1257	9	US-10-074-095-996	Sequence 996, App
30	38	1.6	1257	10	US-09-764-860-996	Sequence 996, App
31	37.8	1.6	544	9	US-10-198-846-13030	Sequence 13030, A
32	37.8	1.6	852	9	US-10-198-846-6037	Sequence 6037, Ap
33	37.8	1.6	3831	12	US-10-051-952-5	Sequence 5, App1
34	37.4	1.6	8761	9	US-10-098-841-11	Sequence 11, App1
35	37.4	1.6	10758	12	US-10-044-090-61	Sequence 61, App1
36	37.4	1.6	1830121	9	US-10-329-860-1	Sequence 1, App1
37	37.2	1.5	606	9	US-10-106-698-208	Sequence 208, App
38	37.2	1.5	606	10	US-09-925-301-120	Sequence 120, App
39	37.2	1.5	3996	9	US-10-087-464-42	Sequence 42, App1
40	37	1.5	1799	7	US-08-781-986A-447	Sequence 447, App
41	36.8	1.5	7153	9	US-10-239-676-69	Sequence 69, App1
42	36.6	1.5	7101	9	US-10-011-366-9	Sequence 9, App1
43	36.4	1.5	756	9	US-10-244-292-6	Sequence 6, App1
44	36.4	1.5	17142	9	US-10-239-676-305	Sequence 205, App
45	36.2	1.5	715	9	US-10-198-846-336	Sequence 336, App

## ALIGNMENTS

RESULT 1  
US-09-986-682B-5  
Sequence 5, Application US/09986682B  
Patent No. US20020115182A1  
GENERAL INFORMATION:  
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU  
KENKYUJO  
TITLE OF INVENTION: POLYPEPTIDE HAVING - FRUCTOFURANOSIDASE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/986, 682B  
FILING DATE: 06-No. US20020115182A1-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 170, 630/1996  
FILING DATE: 10-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25, 618  
REFERENCE/DOCKET NUMBER: TSUSAKI-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-737-3528  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2408 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Bacillus sp.  
INDIVIDUAL ISOLATE: V230 (FERM BP-5054)  
FEATURE:  
NAME/KEY: signal peptide  
LOCATION: 1..360  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
us-09-986-682b-5

Query Match 100.0% Score 2408; DB 10; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGGGAAATCTGATTCATTCGATGCGAGAGCTTCCCGATGGTGTAGAGAGAGTTCG 60  
DB 1 CGGGGAAATCTGATTCATTCGATGCGAGAGCTTCCCGATGGTGTAGAGAGAGTTCG 60  
OY 61 GACTGCCAATGCAGCTGTGCGTAAGAAACAGCTTACTCATGAGCAATTACTAGAGAAT 120  
DB 61 GACTGCCAATGCAGCTGTGCGTAAGAAACAGCTTACTCATGAGCAATTACTAGAGAAT 120  
OY 121 TTCAAAATCTCGGATTAAGCCCAAGTCCACTTTGAAATTAACCTTTGAGCCTCTGTGTG 180  
DB 121 TTCAAAATCTCGGATTAAGCCCAAGTCCACTTTGAAATTAACCTTTGAGCCTCTGTGTG 180  
OY 181 GGGGCTTTTGTGTTTATTTATTCACAGTGCAGTGGTCCATCCCTATATTCATTTAA 240  
DB 181 GGGGCTTTTGTGTTTATTTATTCACAGTGCAGTGGTCCATCCCTATATTCATTTAA 240  
OY 241 GACGAAATTCATCAATCATGCCATCCCAATTAACCTGCTCTCTATCTTTTAA 300  
DB 241 GACGAAATTCATCAATCATGCCATCCCAATTAACCTGCTCTCTATCTTTTAA 300  
OY 301 TTAATAAACAATATCAGAGCTTTCTTATCAATTCATACATATCCAGAGAGAGAGC 360  
DB 301 TTAATAAACAATATCAGAGCTTTCTTATCAATTCATACATATCCAGAGAGAGAGC 360  
OY 361 ATGAACCTTCAAAAGATTGGCGAAAAAAGACAGTGGCGTAACCTTCAGAGTGTATTTA 420  
DB 361 ATGAACCTTCAAAAGATTGGCGAAAAAAGACAGTGGCGTAACCTTCAGAGTGTATTTA 420  
OY 421 GTAGAGCGGAGCGAGCCGATATTTTTCGCGAGCAATGAACAGCGGGGACTTACAGGAA 480  
DB 421 GTAGAGCGGAGCGAGCCGATATTTTTCGCGAGCAATGAACAGCGGGGACTTACAGGAA 480  
OY 481 GACTATGTTTGGCCATATTAACAGCGGCTGACATGCTTAAATTCAGAGACAAAC 540  
DB 481 GACTATGTTTGGCCATATTAACAGCGGCTGACATGCTTAAATTCAGAGACAAAC 540  
OY 541 AGTCCCTAATTTAAAGTCCCTCAATTCATGATCAGCAATCAAAAACATTGATTCGCA 600  
DB 541 AGTCCCTAATTTAAAGTCCCTCAATTCATGATCAGCAATCAAAAACATTGATTCGCA 600  
OY 601 AAAGGATGTGAATAGTCAAGCACTTAATAGATTAGATGATGAGGATAGCTGGCCACTG 660  
DB 601 AAAGGATGTGAATAGTCAAGCACTTAATAGATTAGATGATGAGGATAGCTGGCCACTG 660  
OY 661 CAAAACGCTGATGAGTACGCGGCAATTAATCATGATATCAATCAGTCTCGCTTTTACA 720  
DB 661 CAAAACGCTGATGAGTACGCGGCAATTAATCATGATATCAATCAGTCTCGCTTTTACA 720  
OY 721 GGTGACCCAAAACAGTATGATATCTCACTTCTATTCATCAAAAAGTGGGTAT 780  
DB 721 GGTGACCCAAAACAGTATGATATCTCACTTCTATTCATCAAAAAGTGGGTAT 780  
OY 781 ACATCGATTGACAGCTGGAAAAATGCTGGAAGAGATTGTAAGATATGATATTAATTTGTT 840  
DB 781 ACATCGATTGACAGCTGGAAAAATGCTGGAAGAGATTGTAAGATATGATATTAATTTGTT 840  
OY 841 CCAATATGATCGGTATCTTAATATCAACACAGAGATGTGAGTTCGTCTACTTTAAC 900  
DB 841 CCAATATGATCGGTATCTTAATATCAACACAGAGATGTGAGTTCGTCTACTTTAAC 900

DB 841 CCAATATGATCGGTATCTTAATATCAACACAGAGATGTGAGTTCGTCTACTTTAAC 900  
OY 901 AAAGATGGCCCAAGTCCGTTTATTTCTATACATATTCAGGATATCTCTGAAGATGGTGA 960  
DB 901 AAAGATGGCCCAAGTCCGTTTATTTCTATACATATTCAGGATATCTCTGAAGATGGTGA 960  
OY 961 ACCGCTGTGTAAACCAATCATTTCAACTGCTCAAGTAAATATCCACGCCGATGCA 1020  
DB 961 ACCGCTGTGTAAACCAATCATTTCAACTGCTCAAGTAAATATCCACGCCGATGCA 1020  
OY 1021 GCTACACTTAAAGTGCATGAGATCTGATCATTAATCTGTTTGTATGGCGAGACGT 1080  
DB 1021 GCTACACTTAAAGTGCATGAGATCTGATCATTAATCTGTTTGTATGGCGAGACGT 1080  
OY 1081 ACAGTTTATCAAAATATTCACAGCAATTAATCATGAAGCAAGTGGATTTCAGGTATAC 1140  
DB 1081 ACAGTTTATCAAAATATTCACAGCAATTAATCATGAAGCAAGTGGATTTCAGGTATAC 1140  
OY 1141 CATACCTTAAAGAGACCTCACTATGTTGAAGATGAAGGCCATAATATCTTGTGAA 1200  
DB 1141 CATACCTTAAAGAGACCTCACTATGTTGAAGATGAAGGCCATAATATCTTGTGAA 1200  
OY 1201 GCGAATATCGAACAACAGATGTTATCAAGCGCATAGTCTTCAATTAATAAAGCTTAC 1260  
DB 1201 GCGAATATCGAACAACAGATGTTATCAAGCGCATAGTCTTCAATTAATAAAGCTTAC 1260  
OY 1261 TATGCGGAAATGACGCTCTTCTTCAGAAATGAATAAATACGTTCAAAAGCTTAA 1320  
DB 1261 TATGCGGAAATGACGCTCTTCTTCAGAAATGAATAAATACGTTCAAAAGCTTAA 1320  
OY 1321 AAACAAATGCTTCTTTAGCGAATGATGATTAAGCATTTGTTGAATTTGGCCGATAT 1380  
DB 1321 AAACAAATGCTTCTTTAGCGAATGATGATTAAGCATTTGTTGAATTTGGCCGATAT 1380  
OY 1381 ACAATGAAATATGTTATGAACCAATAGTCCATCAAAACAGATGAAGTGA 1440  
DB 1381 ACAATGAAATATGTTATGAACCAATAGTCCATCAAAACAGATGAAGTGA 1440  
OY 1441 CGCGCAATATATTTAAATGAATGAATGAATGATGATGATGATGATGATGATGATGAT 1500  
DB 1441 CGCGCAATATATTTAAATGAATGAATGAATGATGATGATGATGATGATGATGATGAT 1500  
OY 1501 AAATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
DB 1501 AAATGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
OY 1561 TCCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
DB 1561 TCCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
OY 1621 GACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
DB 1621 GACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
OY 1681 AATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
DB 1681 AATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
OY 1741 CTGGGGAACAGCTGGGGTTAATATTAAGGCTGACACATCTGGAGGAAATATGAT 1800  
DB 1741 CTGGGGAACAGCTGGGGTTAATATTAAGGCTGACACATCTGGAGGAAATATGAT 1800  
OY 1801 TCCGAGACAGCAATTCCTATGAGATATCTCCCATTAATAAAGATATGCTACCTGCA 1860  
DB 1801 TCCGAGACAGCAATTCCTATGAGATATCTCCCATTAATAAAGATATGCTACCTGCA 1860  
OY 1861 GATCTACCTTTTCCCAACCTTTTAAATTAAGGTTTGAACACCTTTGAA 1920  
DB 1861 GATCTACCTTTTCCCAACCTTTTAAATTAAGGTTTGAACACCTTTGAA 1920  
OY 1921 GAAAAAAGATCTCTGGCCCGGCAATATCCCATGCGAAGAGCTACCCCAATAA 1980  
DB 1921 GAAAAAAGATCTCTGGCCCGGCAATATCCCATGCGAAGAGCTACCCCAATAA 1980

QY 1981 AGAATTTTCCCTTTTATAGGGGGGAGTCCCTTTTCTATCTTGGGGTGGGATTTGTG 2040  
D 1981 AGAATTTTCCCTTTTATAGGGGGGAGTCCCTTTTCTATCTTGGGGTGGGATTTGTG 2040  
QY 2041 GCGCCACACGAGGACCTTTATTTATTTATGATGATCTGCGATTTTGGTGGCCTTA 2100  
D 2041 GCGCCACACGAGGACCTTTATTTATTTATGATGATCTGCGATTTTGGTGGCCTTA 2100  
QY 2101 CTTATTTACAGGCGGCGCCGCAAAATTTAATACCAAGAACTCTTAAGTATGGCGGT 2160  
D 2101 CTTATTTACAGGCGGCGCCGCAAAATTTAATACCAAGAACTCTTAAGTATGGCGGT 2160  
QY 2161 ACAATTTACCAACACCTGCTGAGAGGAGTATCCGCCCTCAATACCACTTTCTGTTC 2220  
D 2161 ACAATTTACCAACACCTGCTGAGAGGAGTATCCGCCCTCAATACCACTTTCTGTTC 2220  
QY 2221 CCGCAAAATGGAATAATGATCCGAGAGGCCCATCTTTTGGGGGAAGTATCATTTACTA 2280  
D 2221 CCGCAAAATGGAATAATGATCCGAGAGGCCCATCTTTTGGGGGAAGTATCATTTACTA 2280  
QY 2281 TTACCTTTACCAAGAACTATCCAGATGGAATGTAAGGATGGGCGCATGCAAGCTC 2340  
D 2281 TTACCTTTACCAAGAACTATCCAGATGGAATGTAAGGATGGGCGCATGCAAGCTC 2340  
QY 2341 CGAAGATTTTATGCTGAGAGGAGGAGGATTCCTCCGAGATATACCAATTAATAA 2400  
D 2341 CGAAGATTTTATGCTGAGAGGAGGAGGATTCCTCCGAGATATACCAATTAATAA 2400  
QY 2401 TGGTGATC 2408  
D 2401 TGGTGATC 2408

RESULT 2

US-09-986-682b-4  
Sequence 4, Application US/09986682B  
Patent No. US20020115182A1

GENERAL INFORMATION:

APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU  
KENTYUHO

TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE  
ACTIVITY

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/986,682B

FILING DATE: 06-No. US20020115182A1-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 170,630/1996

FILING DATE: 10-JUN-1996

ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TSUSAKI-2

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1365 base pairs

TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-986-682b-4

Query Match 56.7%; Score 1365; DB 10; Length 1365;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ATGAACAGCGGGGACTACAGAGAACTATGTTTGGCCATATTACAGCGCTGACATG 516  
D 1 ATGAACAGCGGGGACTACAGAGAACTATGTTTGGCCATATTACAGCGCTGACATG 60  
QY 517 CTAATAATTCAGGACCAACAACAGTCCCAATTTAAGGCGCTCAATTCATGACATCA 576  
D 61 CTAATAATTCAGGACCAACAACAGTCCCAATTTAAGGCGCTCAATTCATGACATCA 120  
QY 577 GCAATCAAAAACATTTGATTCGGCAAAAGGATATGATAGTACAGCACTTAATGATTTA 636  
D 121 GCAATCAAAAACATTTGATTCGGCAAAAGGATATGATAGTACAGCACTTAATGATTTA 180  
QY 637 GATGATGGGATPAGCTGGCCACTGCAAAACGCTGATGCTGCGGCAATTTATCARGA 696  
D 181 GATGATGGGATPAGCTGGCCACTGCAAAACGCTGATGCTGCGGCAATTTATCARGA 240  
QY 697 TATCAGATCGTCTCGCTTTAGCAGGTGACCCCAAAAACAGTGAATGATCTCCACTTCAT 756  
D 241 TATCAGATCGTCTCGCTTTAGCAGGTGACCCCAAAAACAGTGAATGATCTCCACTTCAT 300  
QY 757 TTATTTATCAAAAAGTCTGATATACATGATTCAGACCTGGAATAATGCTGAAGAGTA 816  
D 301 TTATTTATCAAAAAGTCTGATATACATGATTCAGACCTGGAATAATGCTGAAGAGTA 360  
QY 817 TTTGAGATATGATTAATTTGTTCCAAATGATATCCGATCTTAATATATCAACACAGAG 876  
D 361 TTTGAGATATGATTAATTTGTTCCAAATGATATCCGATCTTAATATATCAACACAGAG 420  
QY 877 TGGTCAGGTTCTGCTACTTTAACAAGATGGCCAGTCCGTTATTTCTATACAGATTTAC 936  
D 421 TGGTCAGGTTCTGCTACTTTAACAAGATGGCCAGTCCGTTATTTCTATACAGATTTAC 480  
QY 937 TCAGTAATCTCGAAGATGATGTAACCGGCTGCTGTAACCAATCATTTCACTGCTCAA 996  
D 481 TCAGTAATCTCGAAGATGATGTAACCGGCTGCTGTAACCAATCATTTCACTGCTCAA 540  
QY 997 GTAACCTTATCCACCGCGGATGACGCTACCTTAAAGTCGATGATCATPAA 1056  
D 541 GTAACCTTATCCACCGCGGATGACGCTACCTTAAAGTCGATGATCATPAA 600  
QY 1057 TCTGCTTTGATGGCGGAGAGGTAGCAAGTTATCAAAATATTCAGCAATTTATGATGAA 1116  
D 601 TCTGCTTTGATGGCGGAGAGGTAGCAAGTTATCAAAATATTCAGCAATTTATGATGAA 660  
QY 1117 GCGAAGTGAATTTAGGATTAACCATCTTTAAGAGCCCTCATGTTGAAGATPAG 1176  
D 661 GCGAAGTGAATTTAGGATTAACCATCTTTAAGAGCCCTCATGTTGAAGATPAG 720  
QY 1177 GCGCATTAATATCTTGTCTTTGAAGCGAATACGGAACACAGATGTTATCAAGCGCAT 1236  
D 721 GCGCATTAATATCTTGTCTTTGAAGCGAATACGGAACACAGATGTTATCAAGCGCAT 780  
QY 1237 CAGCTTTTCATTAATAAGCTTACTATGGCGGAGTACGCTCTTCCAGAAATGAATAA 1296  
D 781 CAGCTTTTCATTAATAAGCTTACTATGGCGGAGTACGCTCTTCCAGAAATGAATAA 840  
QY 1297 AATAAAGCTGTTCAAGTCTTAATAAACAATTTGCTTTAGGGAATGATGATGATG 1356  
D 841 AATAAAGCTGTTCAAGTCTTAATAAACAATTTGCTTTAGGGAATGATGATGATG 900  
QY 1357 ATTGTTGATTTGGCGGATGACTATACAGTGAAGAAGTGTATGAACCATTTAGTGCATCA 1416  
D 901 ATTGTTGATTTGGCGGATGACTATACAGTGAAGAAGTGTATGAACCATTTAGTGCATCA 960





1681 AATGTGACTACAGAGTTATATGACGATAGAGGCTTCTATCCAGAACTACCTCTAC 1740  
1294 GAATCGCATTAACAGATACACAAAGAGGAGATTTGAGACGACGATCAGCCACG 1353  
1741 CTGGGGGACAGCTTGGGGTTAATATTAAGGGTCTGACATCTGGAGGAGAAATAG 1799  
1354 TTTCACCAACGCTTTTGGCTGAGATCAAAAGATCAAAACATCCGTTGTCAAAAACAG 1412

RESULT 4  
US-10-055-794-1  
; Sequence 1, Application US/10055794  
; Publication No. US20030059900A1  
; GENERAL INFORMATION:  
; APPLICANT: Farmer, Alan Andrew  
; TITLE OF INVENTION: SEQUENCE SPECIFIC RECOMBINASE-BASED  
; TITLE OF INVENTION: METHODS FOR PRODUCING INTRON CONTAINING VECTORS AND  
; FILE REFERENCE: CION-069  
; CURRENT FILING DATE: 2002-01-17  
; PRIOR APPLICATION NUMBER: 60/263,358  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 4938  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: Synthetic Vector  
US-10-055-794-1

Query Match 25.3%; Score 608.8; DB 9; Length 4938;  
Best Local Similarity 66.4%; Pred. No. 1.3e-152;  
Matches 994; Conservative 0; Mismatches 447; Indels 57; Gaps 6;

358 ACGATGAACCTTCAGATGTTGGCGAAAAAGACAGTGGCGTACCTTCAGAGCTGCTATA 417  
1506 ACGATGAACCTTCAGATGTTGGCGAAAAAGACAGTGGCGTACCTTCAGAGCTGCTATA 1565  
418 TTATAGAGAGGAGGAGCGACCGCATATTTTTCGCGAAGCAATGAACGGCGGACTACAG 477  
1566 CTGCGAGAGAGG---CGCAACTCAAGCGTTTTCGGAAGAAAGCAACCAATGATATAG 1622  
478 GAAACATATGCTTTTGGCCATATTAACGCGCTGACATGCTTAATAATTCAGGACACAA 537  
1623 GAAACATATGCTTTTGGCCATATTAACGCGCTGACATGCTTAATAATTCAGGACACAA 1682  
538 AACAGCTCCTCAATTTAAAGTGCCTCAATTCATGATGATGATGATGATGATGATGATG 597  
1683 AAAAATGAAAATATCAAGTTCCTGAGTTCATGCTGCTCAATTAATAATATCTCTCT 1742  
598 GCAAAAGGATATGATTAAGTACAGCAACTTAATAGATTAGATGATGATGATGATGATG 657  
1743 GCAAAAGG-----CTGAGCGTTTGGGACAGCTGGCCA 1775  
658 CTGCAAAAAGCTGATGCTACTGCGGCAATTAATCATGATATCATGCTCTCGCTTTA 717  
1776 TTACAAAAGGCTGAGCGGCACTGTCGCAAACTACCGCTACCAATGCTCTTGGCATTA 1835  
718 GCAGGTGACCCCAAAAACAGTATGATCTGCACTCTTATTTATCTATCAAAAAGTGGT 777  
1836 GCCGAGATCTCTAAAAATGCGGATGACACATCGATTTTATCATGTTCTTATCAAAAAGTGGC 1895  
778 GATACATCATGATGACAGCTGAAAAATGCTGAAAGATATTGGAAGATATGATTAATTT 837  
1896 GAACTTCTATTTGACAGCTGAAAAAGCTGCGCGCTTTTAAAGACAGGCAAAATTC 1955  
838 GTTCCAATGATCGTATCTTAATATTAACACAGGAGTGTGCTAGTTCTGCTACTTTA 897  
1956 GATGCAATGATCTATCTTAATAAGACCAACACAAAGATGCTGAGTTGAGCCCATTT 2015

898 ACCAAGATGCGCAAGTCCGTTTATCTATACAGATTACTCAGGTAATCCGTAAGATGCT 957  
2016 ACATCTGACGGAATAATCCGTTTATCTACACTGATTTTCCGGTAAACATTA----- 2068  
958 GGAACGGTGTGCTTAACCAATCATTTCAACTGCTCAGTAATACTTAATCCACCGGAT 1017  
2069 -----CGGCAACAAACACTGACACTGACACAGTAAGTAAAGTACGATCAGATCAGAC 2117  
1018 GCAGCTACCTTAAAGTGCAGTATCTATCATTAATCTGCTTTGATGGGGAGAC 1077  
2118 AGCTCT---TTGAACATCAACGGTGTAGAGATTAATAATCAATCTTTGA---CGGTGAC 2171  
1078 GGTACAGTTTATCAAAATATTCAGCAATTTATGATGATGATGATGATGATGATGATGAT 1137  
2172 GGAATAAGCTATCAAAATGCTACAGAGTTCATCATGATGATGATGATGATGATGATGAT 2231  
1138 AACCATCTTTAAGAGACCTTCACTATGTTGAAGTAAGGCGCATTAATATCTGCTCTT 1197  
2232 AACCATGAGTGAAGATCTCATACGCTGAGAGTAAGTAAAGCCCAAAATCTTATGATTT 2291  
1198 GAAGCATATCTGAGACACAGATGTTATGCAAGGCGATCAGTCTTCAATTAATAAGCT 1257  
2292 GAAGCAAACTGAGTGAAGTGAAGTGTCCCAAGGCGAAGATCTTTATTAACAAACCA 2351  
1258 TACTATGCGGAGAGTGAAGTCTTCTTCCAGATGAAAAAATAAATCTGCTTCAAGTCT 1317  
2352 TACTATGCGGAGAGTGAAGTCTTCTTCCAGATGAAAAAATAAATCTGCTTCAAGTCT 2411  
1318 AAAAAACAATTCCTTTTAAAGCAATGATGATGATGATGATGATGATGATGATGATGAT 1377  
2412 AAAAAACGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2471  
1378 TATACAGTGAAGAGTGTATGAAACCATTAAGTATGATGATGATGATGATGATGATGAT 1437  
2472 TACACACTGAAAAAGTATGAAACCGCTGATGATGATGATGATGATGATGATGATGAT 2531  
1438 GAAGCGGCAATATTTTAAATGATATTAATGATGATGATGATGATGATGATGATGATGAT 1497  
2532 GAAGCGGCAATATTTTAAATGATATTAATGATGATGATGATGATGATGATGATGATGAT 2591  
1498 TCCAAAATGACAGTATGATGATTAACGACAAAGATTTTATGATGATGATGATGATGAT 1557  
2592 TCCAAAATGACAGTATGATGATTAACGACAAAGATTTTATGATGATGATGATGATGAT 2651  
1558 GATCTCTTAATGCGCCACACAAAGGATTAATGAACTGATGATGATGATGATGATGATGAT 1617  
2652 AATCTTTTAACTGCGCATTAACAGCGGCTGAACAAAGCTGCTGCTGCTGCTGCTGCT 2711  
1618 CTGACCGCTGATCTACACACACTACTCTCTCTGATGATGATGATGATGATGATGATGAT 1677  
2712 CTGATCTTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2771  
1678 AATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737  
2772 AACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2831  
1738 CACCTGCGGACAGACTTGGGTTATATTAAGGCTGACACATCTGAGAGGAGAAAT 1797  
2832 AGTTTGGCGCTAGCTCTCTGTAACATCAAGGCAAGAAACATCTGTTGTTGTAAGAC 2891  
1798 AGTTCC---GGACAGAGCAATTCCTCATAGGATATCTCCCAATTAATAAGAAATGTC 1852  
2892 AGCATCTTGAACAGACAAATTAACAGTAAACAATTAATAAGCAAGAAATATGCC 2949

RESULT 5  
US-10-055-794-3  
; Sequence 3, Application US/10055794  
; Publication No. US20030059900A1  
; GENERAL INFORMATION:  
; APPLICANT: Farmer, Alan Andrew  
; TITLE OF INVENTION: SEQUENCE SPECIFIC RECOMBINASE-BASED  
; TITLE OF INVENTION: METHODS FOR PRODUCING INTRON CONTAINING VECTORS AND



NAME: Daniel Hart  
 REGISTRATION NUMBER: 40,637  
 REFERENCE/DOCKET NUMBER: GENSET.15CDV1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10317 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: DOUBLE  
 TOPOLOGY: CIRCULAR  
 MOLECULE TYPE: synthetic DNA  
 ORIGINAL SOURCE:  
 ORGANISM: Cloning vector pGendL  
 FEATURE:  
 NAME/KEY: pGendL  
 LOCATION: 1..10317  
 FEATURE:  
 NAME/KEY: Homology with X06404 compl (411..1668)  
 LOCATION: 9..1266  
 IDENTIFICATION METHOD: blastn against X06404  
 FEATURE:  
 NAME/KEY: Kanamycin resistance gene CDS  
 LOCATION: 142..957  
 IDENTIFICATION METHOD: By homology to X06404  
 FEATURE:  
 NAME/KEY: Tn1000's right end  
 LOCATION: complement 1332..1371  
 IDENTIFICATION METHOD: blastn against X60200)  
 FEATURE:  
 NAME/KEY: Homology with U46017 (1-472)  
 LOCATION: 1423..1894  
 IDENTIFICATION METHOD: blastn against U46017  
 FEATURE:  
 NAME/KEY: single stranded DNA replication origin  
 LOCATION: 1423..1894  
 IDENTIFICATION METHOD: By homology to U46017  
 OTHER INFORMATION: mutation T -> C 1658  
 FEATURE:  
 NAME/KEY: Homology with U51113 (2382..6997)  
 LOCATION: 1896..6544  
 IDENTIFICATION METHOD: blastn against U51113  
 FEATURE:  
 NAME/KEY: OriS  
 LOCATION: 1972..2188  
 IDENTIFICATION METHOD: By homology to U51113  
 FEATURE:  
 NAME/KEY: repELR  
 LOCATION: 2897..2918  
 OTHER INFORMATION: Described in seqid 16  
 FEATURE:  
 NAME/KEY: RepE  
 LOCATION: 2903..3034  
 IDENTIFICATION METHOD: By homology to U51113  
 FEATURE:  
 NAME/KEY: T3  
 LOCATION: 3043..3059  
 OTHER INFORMATION: Described in seqid 17  
 FEATURE:  
 NAME/KEY: LRT3RA  
 LOCATION: complement 3045..3069  
 OTHER INFORMATION: Described in seqid 15  
 FEATURE:  
 NAME/KEY: Incc  
 LOCATION: 3070..3320  
 IDENTIFICATION METHOD: By homology to U51113  
 OTHER INFORMATION: Insertion 33 bases 3038..3071  
 FEATURE:  
 NAME/KEY: Para  
 LOCATION: 3655..4821  
 IDENTIFICATION METHOD: By homology to U51113  
 OTHER INFORMATION: mutation G -> A 3878

FEATURE:  
 NAME/KEY: Para  
 LOCATION: 4821..5792  
 IDENTIFICATION METHOD: By homology to U51113  
 FEATURE:  
 NAME/KEY: ParC  
 LOCATION: 5865..6382  
 IDENTIFICATION METHOD: By homology to U51113  
 FEATURE:  
 NAME/KEY: Homology with J01688 (complement 175..819)  
 LOCATION: 6574..7218  
 IDENTIFICATION METHOD: blastn against J01688  
 OTHER INFORMATION: mutation A -> G 7096  
 FEATURE:  
 NAME/KEY: CDS streptomycin sensitivity gene  
 LOCATION: complement 6716..7090  
 IDENTIFICATION METHOD: By homology to J01688  
 OTHER INFORMATION: mutation A -> G 6728  
 mutation G -> C 6821  
 mutation C -> T 6866  
 mutation T -> C 7013  
 mutation T -> A 7058  
 FEATURE:  
 NAME/KEY: tpsLR  
 LOCATION: 7155..7174  
 OTHER INFORMATION: Described in seqid 12  
 FEATURE:  
 NAME/KEY: SP6  
 LOCATION: 7230..7248  
 OTHER INFORMATION: Described in seqid 13  
 FEATURE:  
 NAME/KEY: Tn1000's left end  
 LOCATION: 7252..7291  
 IDENTIFICATION METHOD: blast (X60200)  
 FEATURE:  
 NAME/KEY: Homology with X02730 (complement 37..1959)  
 LOCATION: 7305..9227  
 IDENTIFICATION METHOD: blastn against X02730  
 FEATURE:  
 NAME/KEY: CDS levansucrase gene  
 LOCATION: complement 7379..8800  
 IDENTIFICATION METHOD: By homology to X02730  
 OTHER INFORMATION: mutation T -> C 7466  
 mutation A -> G 7739  
 mutation T -> C (Asn -> Asp) 8347  
 mutation T -> C 8600  
 mutation G -> A (Ala -> Val) 8772  
 FEATURE:  
 NAME/KEY: SLR3  
 LOCATION: 8711..8731  
 OTHER INFORMATION: Described in seqid 14  
 FEATURE:  
 NAME/KEY: Homology with J01636 (complement 1158..1465)  
 LOCATION: 9298..9623  
 IDENTIFICATION METHOD: blastn against J01636  
 FEATURE:  
 NAME/KEY: CDS alpha peptide beta-galactosidase  
 LOCATION: complement 9276..9497  
 IDENTIFICATION METHOD: By homology to J01636  
 FEATURE:  
 NAME/KEY: primer HE1  
 LOCATION: complement 9465..9479  
 FEATURE:  
 NAME/KEY: primer HE2  
 LOCATION: 9461..9475  
 FEATURE:  
 NAME/KEY: primer lacRS2AVT  
 LOCATION: complement 9603..9630  
 FEATURE:  
 NAME/KEY: primer lacE2Mlu  
 LOCATION: 9289..9314  
 FEATURE:  
 NAME/KEY: Homology with M77789 (1889..2576)

LOCATION: 9629..10315  
IDENTIFICATION METHOD: blastn against M77789  
FEATURE:  
NAME/KEY: high copy-number double-stranded DNA replication origin  
IDENTIFICATION METHOD: By homology to M77789  
OTHER INFORMATION: mutation C -> T 9803  
site Scat 10029 - 10034  
CLONING SITES 10031 - 10041  
FEATURE:  
NAME/KEY: oriLRd  
LOCATION: 9856..9881  
OTHER INFORMATION: Described in seqid 8  
FEATURE:  
NAME/KEY: OSI  
LOCATION: 10009..10026  
OTHER INFORMATION: Described in seqid 10  
FEATURE:  
NAME/KEY: ORI  
LOCATION: complement 10046..10062  
OTHER INFORMATION: Described in seqid 11  
FEATURE:  
NAME/KEY: oriLRt  
LOCATION: complement 10182..10202  
OTHER INFORMATION: Described in seqid 9  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-849-866-1

Query Match 25.2% Score 607.2; DB 10; Length 10317;  
Best Local Similarity 66.3%; Pred. No. 5,5e-152;

Matches 993; Conservative 0; Mismatches 448; Indels 57; Gaps 6;

358 ACAGTGAACCTTAAGAATGCGGAAAGAGCGCCCTACCTCGAGCTGCTATA 417  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8803 ACAGTGAACATTAAGAATGCGGAAAGAGCGCCCTACCTCGAGCTGCTATA 417  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
418 TTAGTAGAGCGGAGCGAGCGGATATTTTGGCAGCAATGACAGCGGGGACTACAG 477  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8743 CTGGCAGGAGG---CGCAACTCAAGCGTTTGGCAAGAAACAGCAAAAGCATATAG 8687  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
478 GAAAGCTATGTTTGGCCATTTACAGCGCGCTGACATCTTAAATTTCCAGACACAA 537  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8686 GAAACATAGCGGATCTCCATTTACACCCCATGATGCTCAATCCCTGACAGCAA 8627  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
538 AACAGTCCATTTAAAGTGCCTCAATTCATGATGATGATGATGATGATGATGATG 597  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8626 AAAAATGAATAATCAAGTTCTGAGTTGCTGCTACCAATTTAAATATCTCTTCT 8567  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
598 GCAAAAGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8566 GCAAAAGG-----CCTGAGCGTTTGGGACAGCTGGCCA 8534  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
658 CTGCAAAAGCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 717  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8533 TTACAAAGCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 8474  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
718 GCAGGTGACCAAAAGAGTATGATGATGATGATGATGATGATGATGATGATGATG 777  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8473 GCGGAGATCTTAATAATGCGATGATGATGATGATGATGATGATGATGATGATG 8414  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
778 GATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 837  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8413 GAACTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8354  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
838 GTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 897  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8353 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8294  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
898 ACCAAGATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 957  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8293 ACATCTGACGAAATGCGTTTATCTACACTGATTTCTCGGATAAACATTA----- 8241  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

958 GGAACCGGTGCTGTAACCAATCATTTCACTGCTCAGTAATTAATCCAGCGGAT 1017  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8240 -----CGCAACCAACACTGACATGACAGTAAGTATCAGCATCAAC 8192  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1018 GCAGTACACTTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1077  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8191 AGCTCT---TGACATCAAGGATGATGATGATGATGATGATGATGATGATGATGATG 8138  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1078 GGTACAGTTTCAAAATATTCAGCAATTTATGATGATGATGATGATGATGATGATG 1137  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8137 GGAATACGATCAAAATATTCAGCAATTTATGATGATGATGATGATGATGATGATG 8078  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1138 AACCATCTTTAAGAGACCTCAGTATGATGATGATGATGATGATGATGATGATGATG 1197  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8077 AACCATCTTTAAGAGACCTCAGTATGATGATGATGATGATGATGATGATGATGATG 8018  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1198 GAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1257  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
8017 GAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7958  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1258 TACTATGCGGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATG 1317  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7957 TACTATGCGGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATG 7898  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1318 AAAAAACAATTTGCTTTTACGCAATGATGATGATGATGATGATGATGATGATGATG 1377  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7897 AAAAAACAATTTGCTTTTACGCAATGATGATGATGATGATGATGATGATGATGATG 7838  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1378 TATACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1437  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7837 TACACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7778  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1438 GAACGCGGATATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1497  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7777 GAACGCGGATATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATG 7718  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1498 TCCAAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1557  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7717 TCCAAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7658  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1558 GACTCTTAATGAGGCGGACACACCGGATTAATGATGATGATGATGATGATGATGATG 1617  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7657 AATCTTTTAACTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 7598  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1618 CTGACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1677  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7597 CTGATCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7538  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1678 AATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1737  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7537 AACATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7478  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1738 CACCTGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1797  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7477 ACGTTTGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7418  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1798 AGTTCC---GGACAGCAATTTCCATGAGGATGATGATGATGATGATGATGATGATG 1852  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
7417 ACATCTTTGAACAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 7360  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7  
US-09-993-292A-19  
Sequence 19, Application US/09993292A  
Patent No. US20020146430A1  
GENERAL INFORMATION:  
APPLICANT: James E. Galen  
APPLICANT: University of Maryland  
TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF  
FILE REFERENCE: US/09-007A  
CURRENT APPLICATION NUMBER: US/09/993,292A  
CURRENT FILING DATE: 2001-11-23

: PRIOR APPLICATION NUMBER: 60/252,516  
 : PRIOR FILING DATE: 2000-11-22  
 : NUMBER OF SEQ ID NOS: 19  
 : SOFTWARE: FastSeq for Windows Version 4.0  
 : SEQ ID NO: 19  
 : LENGTH: 2253  
 : TYPE: DNA  
 : ORGANISM: Artificial Sequence  
 : FEATURE:  
 : OTHER INFORMATION: cIYA::SacB fusion gene  
 : NAME/KEY: CDS  
 : LOCATION: (0)...(2253)  
 : US-09-993-292A-19

Query Match 24.1%: Score 580.6; DB 10; Length 2253;  
 Best Local Similarity 67.3%: Pred. No. 2,9e-145;  
 Matches 905; Conservative 0; Mismatches 389; Indels 51; Gaps 4;

Oy 455 AATGACAGCGGGGACTACAGAAGACTATGTTTGGCCATATTACAGCGCTGACA 514  
 Db 920 AAGCAACCAAGGCATATAGAGAAACATAGCGCATTTCCATTTACAGCCATGATA 979  
 Oy 515 TGCATAAAATTCAGAGACAAACAAACAGTCTCAATTTAAAGTGCCTCAATTCATGCAT 574  
 Db 980 TCGTGCACAAATCCCTGAACAGCAAAAAATGAAAAATATCAAGTTCCCTCAATTCGATTCGT 1039  
 Oy 575 CAGCAATCAAAACATTTGATTCGGCAAAAGGTATGATTAAGTCAGCAACTTAATAGATT 634  
 Db 1040 CCACATTTAAATAATCTCTCTGCAAAAG-----CC 1072  
 Oy 635 TAGATGATGGAGATAGTGGCCACTGCAAAACGCTGATGTTCTGGCAATTTATCATG 694  
 Db 1073 TGGACGTTTGGGACAGCTGGCCATATCAAAACGCTGAGCGCATCTGCCAAACTATCAG 1132  
 Oy 695 GATATCAATAGTCTCCGCTTTAGCAGGTGACCCAAAAACAGTATGATCTCCACTTC 754  
 Db 1133 GCTACCAATCATGTCTTTCGATTCAGCGGAGATCTCTAAAAATGCGGATGACATCATGATT 1192  
 Oy 755 ATTATTTCTATCAAAAAGTCGGTGTATCATGATTCAGAGTGGAAAAATGCTGGAAG 814  
 Db 1193 ACATGTTCTATCAAAAAGTCGGGCAAACTTTATTGACAGCTGGAAAAACGCTGCGCG 1252  
 Oy 815 TATTTAAGATATGATTAATTTGTTCCAAATGATCTTATCTTAAATATCAACACAG 874  
 Db 1253 TCTTTAAGACAGCAAAATTCATGCAATGATTTATCTTAAAGCAACACACAG 1312  
 Oy 875 AGTGTGAGTGTCTGCTACTTTAACCAAGATGGCAAGTCCGTTTATTTCTATACAGATT 934  
 Db 1313 AATGTCAGGTTTCAAGCCACATTTACATCTGACGGAAAAATCCGTTTATTTCTACACTGATT 1372  
 Oy 935 ACTCAGTAATCTGGAAGATGTTGGAACCGGTGCTGTACCAAAATCAATTCACACTGCTC 994  
 Db 1373 TCTCCGTAACATTA-----CGGCAAAACAAACACTGACACTGAC 1414  
 Oy 995 AAGTAACCTATCCAGCGGATGCACTACACTTAAAGTGATGAGATCTGATTCATTA 1054  
 Db 1415 AAGTTAAGCTATCAGCATCAACAGCTCT---TTGAACATCAACGGTGTAGAGATTATA 1471  
 Oy 1055 AATCTGCTTTGAGCGCGAAGCGGTACAGTTTATCAAAATTTACGCAATTTATGATG 1114  
 Db 1472 AATCAATCTTTGA---CGGTAGCGAAAAACGTAATTAAGAGAGTTTCACTGATG 1528  
 Oy 1115 AAGCAAGTGGATTCAGAGTATACCATTTTAAAGAGCCCTCACTACTATGTTGAAGATA 1174  
 Db 1529 AAGGCACTACAGCTCAGCGCAACACATACGCTGAGAGATCTCTACACTAGTAAGATA 1588  
 Oy 1175 AGGCCATAAATATCTGCTTTTGAAGCAATCTGGAACAACAGATGTTATCAAGCG 1234  
 Db 1589 AAGGCCAAATAATCTGATTTTGAAGCAACACTGGAATGAGATGCTTACCAAGCG 1648  
 Oy 1235 ATCAGTCTTCAATAAATTAAGCTTACTATGGCGAAGTGCCTCTCTCCAGATGAAA 1294  
 Db 1649 AAGATCTTTTATTTAACAAGCATCTATGGCAAAAGCAATCATTTCTCCGTAAGAAA 1708

Oy 1295 AAAATAAGCTCTCAAGTCTCTAAAAAAACAATTTCTTTTACGAATGTCATTTAG 1354  
 Db 1709 GTCAAAACCTCTGCAAAACGATATAAAAAAGCAGCGCTGATTTAGCAAAAGCGCTCTCG 1768  
 Oy 1355 GCATTGTTGAATTTGGCCGATGACTATACAGTGAAGATGTTATGAACCATTTAGTCGAT 1414  
 Db 1769 GTATGATTTAGCTTAACGATGATTTACACCTGATAAAAAAGTGAACCGCTGATTTGAT 1828  
 Oy 1415 CAACACAGTACGACATGATGACGCGCCCAATATATTAAATGAATTAATGAT 1474  
 Db 1829 CTAAACACAGTAACAGATGAATTTGAACCGCGCAACGCTTTTAAATGAACGCAATGCT 1888  
 Oy 1475 ATCTATTCACAGGATTCAGAGATCCAAATATACGATGATGATTTATACGCAAGATG 1534  
 Db 1889 ACCTGTTTACGATCTACCTCCCGCGATTAATAATGACATTTAGCGCATTTACGTTACGATA 1948  
 Oy 1535 TTTATATGCTAGAGCCCGAGGCGACTCTTAATATGCCCCACACACACCGATTAATGA 1594  
 Db 1949 TTTACATGCTTGTATGTTATGTTCTTAATTTCTTAATCTGACCATACAAAGCGCTGAACAAA 2008  
 Oy 1595 CTGACCTGTATGTAACATGAATCTTGACCTGCTGATCTCACACACTTACTCTATT 1654  
 Db 2009 CTGGCTTGTGTTTAAAAATGATCTTGATCTTACAGATGATTAACCTTTACTTACTCAGACT 2068  
 Oy 1655 GCGGTATCCCGCAGCTGAGGATTAATGTTGTTACTCTCAAGTTATATGACGAATGAG 1714  
 Db 2069 TCGCTGTACTTACAGAAAGAAACATGCTGATTTACAGCTATATGACAAACAGAG 2128  
 Oy 1715 GCTTCTATTCAGAACATCTCTCACTGCGGAGACAGCTTGGGGTTTAAATTTAAAGGT 1774  
 Db 2129 GATCTACGACAGACAAACATCAACGTTTGGCCCAAGCTTCTGCTGAACATCAAGGCA 2188  
 Oy 1775 CTGACACATCTGAGAGAGAAATAG 1799  
 Db 2189 AGAAAAATCTGTTGTCAAAAGCAG 2213

## RESULT 8

US-09-993-292A-18  
 : Sequence 18, Application US/09993292A  
 : Patent No. US20020146430A1  
 : GENERAL INFORMATION:  
 : APPLICANT: James E. Galen  
 : APPLICANT: University of Maryland  
 : TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF  
 : FILE REFERENCE: US/09-993,292A  
 : CURRENT APPLICATION NUMBER: US/09-993,292A  
 : PRIOR FILING DATE: 2001-11-23  
 : PRIOR APPLICATION NUMBER: 60/252,516  
 : NUMBER OF SEQ ID NOS: 19  
 : SOFTWARE: FastSeq for Windows Version 4.0  
 : SEQ ID NO: 18  
 : LENGTH: 8908  
 : TYPE: DNA  
 : ORGANISM: Artificial Sequence  
 : FEATURE:  
 : OTHER INFORMATION: pSEC84sacB vector  
 : US-09-993-292A-18

Query Match 24.1%: Score 580.6; DB 10; Length 8908;  
 Best Local Similarity 67.3%: Pred. No. 7e-145;  
 Matches 905; Conservative 0; Mismatches 389; Indels 51; Gaps 4;

Oy 455 AATGACAGCGGGGACTACAGAAGACTATGTTTGGCCATATTACAGCGCTGACA 514  
 Db 1435 AAGCAACCAAGGCATATAGAGAAACATAGCGCATTTCCATTTACAGCCATGATA 1494  
 Oy 515 TGCATAAAATTCAGAGACAAACAAACAGTCTCAATTTAAAGTGCCTCAATTCATGCAT 574  
 Db 1495 TCGTGCACAAATCCCTGAACAGCAAAAAATGAAAAATATCAAGTTCTCGAATTCGATTCGT 1554



OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.6  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5  
OTHER INFORMATION: NT HIT: X52988.1, EVALUE 0.00e+00  
OTHER INFORMATION: SWISSPROT HIT: P21130, EVALUE 1.00e-64  
OTHER INFORMATION: EST\_HUMAN HIT: BF219305.1, EVALUE 4.00e-07  
NAME/KEY: unsure  
LOCATION: 332  
NAME/KEY: unsure  
LOCATION: 335  
NAME/KEY: unsure  
LOCATION: 356  
NAME/KEY: unsure  
LOCATION: 357  
NAME/KEY: unsure  
LOCATION: 358  
NAME/KEY: unsure  
LOCATION: 359  
NAME/KEY: unsure  
LOCATION: 360  
NAME/KEY: unsure  
LOCATION: 361  
NAME/KEY: unsure  
LOCATION: 362  
NAME/KEY: unsure  
LOCATION: 363  
NAME/KEY: unsure  
LOCATION: 364  
NAME/KEY: unsure  
LOCATION: 365  
NAME/KEY: unsure  
LOCATION: 366  
NAME/KEY: unsure  
LOCATION: 367  
NAME/KEY: unsure  
LOCATION: 368  
US-09-864-761-21746

Query Match 5.6%; Score 135.6; DB 10; Length 368;  
Best Local Similarity 64.0%; Pred. No. 3.6e-26;  
Matches 235; Conservative 0; Mismatches 105; Indels 27; Gaps 1;  
QY 472 TACAAGAGACTGTTTGGCCCATATTTACAGCGCGCTGACAGCTAAATAATTCACAGA 531  
DB 12 TACAAGAGAGACTGCGGCTCTCATATTTACAGCGCATGATGCTGCAGATCCCTAAA 71  
QY 532 CAACAAGAGACTGCTCATTTAAAGTGGCTCATTCATTCATGATGCAATCAAAAACATT 591  
DB 72 CAGACGCAAGAGAGAAATACAGTGGCTCAATTCGATCAACATTAATAATATTT 131  
QY 592 GATTCGGCAAGAGAGATGATAGTACAGCAACTTAATAGATTAGATGTATGGATAGC 651  
DB 132 GAGTCTGCAAGAGACTT-----GATGTGGGACACG 164  
QY 652 TGGCCATGCAAGAGAGCTGATGCTGCGCAATTAATCATGATATCATGCTGCTGC 711  
DB 165 TGGCCGCTGCAAGAGAGCTGACGGAACACTGATCAACAGCGCTATCAGCTGTGTTT 224  
QY 712 GCTTAGCAGTACCCCAAAAAACAGTATGATCTCAGTTCATTATTCATCAAAAA 771  
DB 225 GCTCTTGGGAGAGAGCCGAAAAAGAGCTGATGACACATCAATCTATATTTATCAAAAG 284  
QY 772 GTGCGATATACATGATGACAGCTGGAATAATGCTGGAAGATATTTGAAGATATGAT 831  
DB 285 GTGCGGCAACATCAATGACAGCTGGAATAACGCGGCGCTGCTTTAAGACAGCGAT 344  
QY 832 AATTTTG 838  
111111

DB 345 AAGTTCG 351  
RESULT 10  
US-09-974-300-735  
Sequence 735, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
APPLICANT: Clausen, Id Groth  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085.500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
PRIOR FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 735  
LENGTH: 882  
TYPE: DNA  
ORGANISM: Bacillus licheniformis  
US-09-974-300-735

Query Match 4.6%; Score 110.4; DB 10; Length 882;  
Best Local Similarity 71.8%; Pred. No. 3.8e-19;  
Matches 158; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
QY 2190 GGAATATCCCGCCCAATACCACCTTCTGTCGCCAAGTGAATAAGATCCGACAGG 2249  
DB 101 GAATATCGTGACGCTTTCATTTGACAACCCCTGACAATGGAATAAGACCTCAAAAA 160  
QY 2250 CCCATCTTTTGGGGA-AGTATCATTAATTAATCACTTTTACAAAGACTATCCAGAT 2308  
DB 161 CCCGCTATTTCATGAGAAATATCATTAATCACTGCTGCTAATACCGAGACTCCGAGC 220  
QY 2309 GGAATGATGAGATGCGGCGCATCAGCAGTCGCAAGATTTATTCATTTGACGAGCAA 2368  
DB 221 GGCATGGAACAGATGGTGGCGCCAGCAGTATCATGATGATCTGCTGCTGCGAGATAAA 280  
QY 2369 GGGATTGCCATCCGAGATATACCAATTAATAATGGTATC 2408  
DB 281 GGGTTGCCATTCGAAATATACGATATAAATGGCATC 320

RESULT 11  
US-09-778-844-136  
Sequence 136, Application US/09778844  
Patent No. US20020150971A1  
GENERAL INFORMATION:  
APPLICANT: JOHANSEN, JEANETTE ELISABETH  
APPLICANT: SCHALLING, MARTIN  
TITLE OF INVENTION: NUCLEIC ACIDS AND POLYPEPTIDES FOR CONTROLLING FOOD  
FILE REFERENCE: 030307/0195  
CURRENT APPLICATION NUMBER: US/09/778,844  
CURRENT FILING DATE: 2001-02-08  
NUMBER OF SEQ ID NOS: 206  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 136  
LENGTH: 1146  
TYPE: DNA  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: CPLA2-beta, EMBL  
US-09-778-844-136

Query Match 3.9%; Score 93; DB 10; Length 1146;  
Best Local Similarity 61.7%; Pred. No. 2.1e-14;





```

: NAME/KEY: CDS
: LOCATION: (1220)..(3598)
: FEATURE:
: NAME/KEY: RBS
: LOCATION: (1205)..(1210)
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: (2702)..(2707)
: OTHER INFORMATION: a, c, t, g, other or unknown
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: (3686)..(3698)
: OTHER INFORMATION: a, c, t, g, other or unknown
US-09-995-587A-10

Query Match
Best Local Similarity 2.5%; Score 61.2; DB 10; Length 4634;
Matches 181; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 1047 TGATCATTAATCTGCTTTGATGCGGAGACGCTACAGTTTATCAAAATATTCAGCAATT 1106
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2299 TGTGATATATATATCAAGTTTGTGTAAGGTGATGATTTCTACCAAACTTATGAAACA 2358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1107 TATGATCAAGGCAAGTGGATTTCAGTGATACCAATCTTAAAGACCTTACTATGT 1166
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2359 ATTGCAACGCGCAAGATCGGAAATGATGATTACTGCTTACGTCGACCCACAGTTGT 2418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1167 TGA---AGATTAAGGCGCATTAATATCTGCTTTGAGGATPACTGGAACAAGATGG 1223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2419 TCAATTAACAAATGCTGATCGTATCTGTATTCAGATTAATGCTGGACAGAAATTA 2478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1224 TTATCAAGGCGATCAGTCTTTCATTAATAAACCTTACTATGAGGGAAGTACGCTTCTT 1283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2479 CCAAGTACGACCAAAATTTTAAATTTGGGCTACATGCTGGCATMGATGCTTCAATAT 2538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1284 CCAGATGAAAAAATAAACGCTTCAAGTCTTAAAAACAATGCTTCTTTAGCGAA 1343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2539 TAAGAGTTCCTCAAGCTTTTGAATAATAGAAGATCCTGAAATGGCTGTTTACCTAA 2598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1344 TGTGCTATTAGGAGCTTTGTAATTTGGCGGATGATATACAGTAAGAAAGTATGAAACC 1403
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2599 TGTGCTACTTGTGATCTTAAAGCTACTTAACATCAAAAGTAAAGCCAAAGTTGAAGAGT 2658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1404 AT 1405
    |||||
Db 2659 AT 2660
    |||||

RESULT 14
US-10-198-846-5829/c
: Sequence 5829, Application US/10198846
: Publication No. US2003009974A1
: GENERAL INFORMATION:
: APPLICANT: Lillie, James
: APPLICANT: Xu, Yougyao
: APPLICANT: Wang, Youzhen
: APPLICANT: Steimann, Kathleen
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
: TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: TITLE OF INVENTION: THERAPY OF BREAST CANCER
: FILE REFERENCE: MRI-049
: CURRENT APPLICATION NUMBER: US/10/198,846
: CURRENT FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/306,220
: PRIOR FILING DATE: 2001-07-18
: NUMBER OF SEQ ID NOS: 14084
: SOFTWARE: fastseq for Windows Version 4.0
: SEQ ID NO 5829
: LENGTH: 839
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature

```

```

: LOCATION: 3, 15, 188, 189, 190, 191, 192, 209, 222, 226, 227, 229,
: LOCATION: 241, 242, 244, 250, 257, 258, 259, 260, 261, 262, 263, 265,
: LOCATION: 281, 285, 291, 292, 293, 294, 295, 296, 298, 303, 305, 307,
: LOCATION: 308, 311, 312, 319, 323, 325, 329, 332, 339, 355, 357
: OTHER INFORMATION: n = A,T,C or G
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 358, 368, 369, 370, 374, 380, 386, 387, 388, 392, 401, 402,
: LOCATION: 403, 406, 409, 413, 414, 415, 426, 427, 439, 453, 460, 465,
: LOCATION: 466, 468, 470, 480, 482, 502, 503, 518, 524, 527, 534, 539,
: LOCATION: 547, 552, 557, 560, 569, 611, 612, 619, 627, 631, 640
: OTHER INFORMATION: n = A,T,C or G
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 644, 651, 658, 659, 672, 675, 676, 677, 679, 681, 691, 693,
: LOCATION: 698, 702, 720, 723, 724, 727, 732, 738, 742, 747, 748, 749,
: LOCATION: 750, 757, 758, 764, 769, 782, 783, 793, 800, 802, 817, 818,
: LOCATION: 821, 826, 828
: OTHER INFORMATION: n = A,T,C or G
US-10-198-846-5829

Query Match
Best Local Similarity 2.2%; Score 52; DB 9; Length 839;
Matches 88; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1882 TTTTATTAATAAGGTTTGTGACCACTTTAGAGAAATAAGAAATCTTGCGCCG 1941
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 TTTTATTTCCCAAGGGGGGTTTTCCTTTTAAAGGAAATAATGNGGGGCG 551
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1942 GGCATTAACCATTCGCCAGAGAGCTACCAATAAAGAAATTTCTTTTAAAGC 2001
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 GGGNCCGCCCGGNGGGGNGGNNTTAAATAATTAATATTAATTTNNCCCCCGGGG 491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2002 GGGGATCTCTTTTCTTCTTCTTGGGTGGGATTTGGCCCCCACC 2049
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 GGGGGCCCTTTTAAANGNNGGGGNTTTTNCCTCCCCC 443
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-09-995-587A-2
: Sequence 2, Application US/09995587A
: Patent No. US20020127681A1
: GENERAL INFORMATION:
: APPLICANT: VAN HIJUM, SACHA ADRIANUS FORKE TACO
: APPLICANT: VAN GEEL-SCHUTTEN, GERRITDIANA HENDRIKA
: APPLICANT: DIJKHUIZEN, LUBBERT
: APPLICANT: RAHOUTI, HAKIM
: TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES
: FILE REFERENCE: BO43667-CIP
: CURRENT APPLICATION NUMBER: US/09/995,587A
: CURRENT FILING DATE: 2002-02-28
: PRIOR APPLICATION NUMBER: 09/604,958
: PRIOR FILING DATE: 2000-06-28
: PRIOR APPLICATION NUMBER: EPO 00201872.9
: PRIOR FILING DATE: 2000-05-25
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 2367
: TYPE: DNA
: ORGANISM: Lactobacillus reuteri
US-09-995-587A-2

Query Match
Best Local Similarity 2.2%; Score 52; DB 10; Length 2367;
Matches 589; Conservative 0; Mismatches 535; Indels 96; Gaps 12;

QY 536 AAACAGTCCCTCAATTTAAAGTCCCATTCATGATCAGCATCAACAACATTGATT 595
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 674 AACAAAGATGTCGGTACACAGTTCATTTTAAAGCAAGTAAATCAAAAATATGCTG 733
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 596 CGGCAAAAGGATATAGTACGAGCACTTAATAGA---TTTAGATGATGAGTACT 652
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 02:52:36 ; Search time 111.688 Seconds  
(Without alignments)  
6611.951 Million cell updates/sec

Title: US-09-986-682b-5

Perfect score: 2408  
Sequence: 1 CCGCGAATAACTAGATTC.....TACCATAAATGTGATC 2408

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata1/ina/5A.COMB.seq: \*  
4: /cgn2\_6/ptodata1/ina/5B.COMB.seq: \*  
5: /cgn2\_6/ptodata1/ina/PCTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2408	100.0	2408	2	US-08-870-827-5
2	2408	100.0	2408	4	US-09-317-179-5
3	1365	56.7	1365	2	US-08-870-827-4
4	1365	56.7	1365	4	US-09-317-179-4
5	607.2	25.2	10317	3	US-09-058-746-1
6	607.2	25.2	10317	4	US-09-438-142-1
7	553.8	23.0	3305	3	US-09-068-043-1
8	44.8	1.9	7218	1	US-08-232-463-14
9	39.8	1.7	1662	1	US-08-671-947-1
10	38	1.6	867	4	US-09-276-531-82
11	37.6	1.5	1434	4	US-09-134-001C-650
12	36.6	1.5	7101	1	US-08-480-604A-9
13	36.6	1.5	7101	2	US-08-405-496A-9
14	36.6	1.5	7101	4	US-08-915-136-9
15	36.6	1.5	7101	4	US-08-957-310-9
16	36.4	1.5	3095	6	5231168-1
17	35.8	1.5	381	4	US-09-134-001C-1861
18	35.4	1.5	849	4	US-09-648-520E-48
19	35.4	1.5	3317	4	US-09-193-562B-1
20	34.8	1.4	6854	4	US-08-961-527-200
21	34.8	1.4	62	3	US-09-068-043-5
22	34.2	1.4	3360	1	US-07-712-833A-1
23	34	1.4	1467	5	US-09-134-001C-1409
24	34	1.4	2817	4	PCT-US93-05944-1
25	34	1.4	3356	4	US-09-379-523-4
26	33.8	1.4	3279	4	US-08-446-137B-1
27	33.8	1.4	51952	3	US-08-947-823-1

C 28	33.6	1.4	1365	2	US-08-870-827-4	Sequence 4, Appl1
C 29	33.6	1.4	1365	4	US-09-317-179-4	Sequence 4, Appl1
C 30	33.6	1.4	2408	2	US-08-870-827-5	Sequence 5, Appl1
C 31	33.6	1.4	2408	4	US-09-317-179-5	Sequence 5, Appl1
C 32	33.6	1.4	3713	1	US-08-100-709-1	Sequence 1, Appl1
C 33	33.6	1.4	3713	1	US-08-176-865-1	Sequence 1, Appl1
C 34	33.6	1.4	3713	1	US-08-474-038-1	Sequence 1, Appl1
C 35	33.6	1.4	3713	2	US-08-779-046-1	Sequence 1, Appl1
C 36	33.6	1.4	3713	2	US-08-881-340-1	Sequence 1, Appl1
C 37	33.4	1.4	1566	4	US-09-134-001C-2802	Sequence 2802, Ap
C 38	33.4	1.4	2022	4	US-08-879-098-1	Sequence 1643, Ap
C 39	33.2	1.4	900	4	US-08-631-548-1	Sequence 1, Appl1
C 40	33.2	1.4	900	4	US-09-631-548-1	Sequence 1, Appl1
C 41	33.2	1.4	3891	1	US-08-480-604A-27	Sequence 27, Appl1
C 42	33.2	1.4	3891	2	US-08-405-496A-27	Sequence 27, Appl1
C 43	33.2	1.4	3891	4	US-08-915-136-27	Sequence 27, Appl1
C 44	33	1.4	699	4	US-08-936-165A-233	Sequence 23, App
C 45	33	1.4	2047	4	US-08-836-261A-1	Sequence 1, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-870-827-5  
Sequence 5, Application US/08870827  
Patent No. 5962297  
GENERAL INFORMATION:  
APPLICANT: Tsusaki et al.  
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE  
TITLE OF INVENTION: ACTIVITY  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/870,827  
FILING DATE: 06-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 170,630/1996  
FILING DATE: 10-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TSUSAKI-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2408 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Bacillus sp.  
INDIVIDUAL ISOLATE: V230 (FERM BP-5054)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..360  
IDENTIFICATION METHOD: E  
NAME/KEY: signal peptide

LOCATION: 361..456  
IDENTIFICATION METHOD: S  
NAME/KEY: mat peptide  
LOCATION: 457..1821  
IDENTIFICATION METHOD: S  
NAME/KEY: 3'UTR  
LOCATION: 1822..2408  
IDENTIFICATION METHOD: E  
US-08-870-827-5

Query Match 100.0%; Score 2408; DB 2; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGGAAAATATCTAGATTCCTCAATGGCCAGACTTCCAGTGGTGAAGAGAGTTCG 60  
DB 1 CGGGGAAAATATCTAGATTCCTCAATGGCCAGACTTCCAGTGGTGAAGAGAGTTCG 60  
QY 61 GACTGCCAATGACAGCTGTGCGTAAGAAAACAGCTTACTCATGAGCAATTAAGAAAT 120  
DB 61 GACTGCCAATGACAGCTGTGCGTAAGAAAACAGCTTACTCATGAGCAATTAAGAAAT 120  
QY 121 TTCAAAGTCTGGGATTAAGGCCCAAGTCCACTTTGAATTAACCTTTGAGCCTGTGTG 180  
DB 121 TTCAAAGTCTGGGATTAAGGCCCAAGTCCACTTTGAATTAACCTTTGAGCCTGTGTG 180  
QY 181 GGGGCTTTTGTATTATTTATTTTCAACTCAAGTGTCCATCCCTATATCAATTTAA 240  
DB 181 GGGGCTTTTGTATTATTTATTTTCAACTCAAGTGTCCATCCCTATATCAATTTAA 240  
QY 241 GAGCAAAATCTAATCAATCCATGCCATCCCAATTAACCTGCTCTCTATACTTTTAA 300  
DB 241 GAGCAAAATCTAATCAATCCATGCCATCCCAATTAACCTGCTCTCTATACTTTTAA 300  
QY 301 TTAATTAAGAACTATCAAGACTTCTTATCAATTCATATATCAAGAGGAGAGAGC 360  
DB 301 TTAATTAAGAACTATCAAGACTTCTTATCAATTCATATATCAAGAGGAGAGAGC 360  
QY 361 ATGAACCTTCAAAAGATTGGCGAAAAAAGCAGCTGCCGTACCTTCAGAGCTGTATTTA 420  
DB 361 ATGAACCTTCAAAAGATTGGCGAAAAAAGCAGCTGCCGTACCTTCAGAGCTGTATTTA 420  
QY 421 GTAGAGGAGGAGACCGCATATTTTGGGAGCAAAATGAACAGCGGGGATACAAAGAA 480  
DB 421 GTAGAGGAGGAGACCGCATATTTTGGGAGCAAAATGAACAGCGGGGATACAAAGAA 480  
QY 481 GACTATGCTTTTGGCCATATTAACAGCGGCTGACATGCTAAATATCCAGCAACAAAC 540  
DB 481 GACTATGCTTTTGGCCATATTAACAGCGGCTGACATGCTAAATATCCAGCAACAAAC 540  
QY 541 AGTCCCAATTTAAAGTGCCTCAATTCATGATCAGCAATCAAAAACATTTGATTCGCA 600  
DB 541 AGTCCCAATTTAAAGTGCCTCAATTCATGATCAGCAATCAAAAACATTTGATTCGCA 600  
QY 601 AAAGGATATGATAGTCAAGCACTTAATTAAGATTTAGATGTATGATGATGCTGCCACTG 660  
DB 601 AAAGGATATGATAGTCAAGCACTTAATTAAGATTTAGATGTATGATGATGCTGCCACTG 660  
QY 661 CAAAACGCTGATGCTGCGCAAAATTTATCATGATATCATGCTGCTGCCCTTTTGA 720  
DB 661 CAAAACGCTGATGCTGCGCAAAATTTATCATGATATCATGCTGCTGCCCTTTTGA 720  
QY 721 GGTGACCCAAAACAGTATGATATCTCACTTATTTATTCAAAAGTGGGAT 780  
DB 721 GGTGACCCAAAACAGTATGATATCTCACTTATTTATTCAAAAGTGGGAT 780  
QY 781 ACATCGATGACAGCTGGAAGAAATGCTGGAAGATTTGAAGATGATGATGATTTGTT 840  
DB 781 ACATCGATGACAGCTGGAAGAAATGCTGGAAGATTTGAAGATGATGATGATTTGTT 840  
QY 841 CCAATATGCGGTATCTTAAATATCAACAGAGAGGTGATGATGCTGCTACTTAAC 900  
DB 841 CCAATATGCGGTATCTTAAATATCAACAGAGAGGTGATGATGCTGCTACTTAAC 900

QY 901 AAAGTGGCCAAAGTCCGTTTATTTCTATACAGATTACTAGATTAATCCGAAGATGTGA 960  
DB 901 AAAGTGGCCAAAGTCCGTTTATTTCTATACAGATTACTAGATTAATCCGAAGATGTGA 960  
QY 961 ACCGCTGTGTGAACCAATATCTTCACTGCTCAAGTAAACTTTATCCAGCCGGATGA 1020  
DB 961 ACCGCTGTGTGAACCAATATCTTCACTGCTCAAGTAAACTTTATCCAGCCGGATGA 1020  
QY 1021 GCTACCTTAAAGTGTGATGATGATATCAATTAATCTGCTTTGATGATGCGGAGAGGT 1080  
DB 1021 GCTACCTTAAAGTGTGATGATGATATCAATTAATCTGCTTTGATGATGCGGAGAGGT 1080  
QY 1081 ACAGTTTATCAAAATATTTACAGCAATTTATGATGATGATGATGATGATGATGATGAT 1140  
DB 1081 ACAGTTTATCAAAATATTTACAGCAATTTATGATGATGATGATGATGATGATGATGAT 1140  
QY 1141 CATACTTTAAGAGACCTCACTATGTTGAAGATTAAGGCGCATAAATATCTTGTCTTGA 1200  
DB 1141 CATACTTTAAGAGACCTCACTATGTTGAAGATTAAGGCGCATAAATATCTTGTCTTGA 1200  
QY 1201 GCGAATCTGGAACACAGATGTTATCAAGGCGATGATGATGATGATGATGATGATGAT 1260  
DB 1201 GCGAATCTGGAACACAGATGTTATCAAGGCGATGATGATGATGATGATGATGATGAT 1260  
QY 1261 TATGCGGAGTGAAGTCTTCTTCCAGAAATGAAGAAATTAACGTTCAAAAGCTTAA 1320  
DB 1261 TATGCGGAGTGAAGTCTTCTTCCAGAAATGAAGAAATTAACGTTCAAAAGCTTAA 1320  
QY 1321 AAACAAATGCTTCTTTAGCGAATGTTGATTAAGCATTTGTTGAATTTGCGCGATGAT 1380  
DB 1321 AAACAAATGCTTCTTTAGCGAATGTTGATTAAGCATTTGTTGAATTTGCGCGATGAT 1380  
QY 1381 ACAGTGAAGTGTATGAAGCAATTTAGTGCATCAAAACAGATGAAGATGAAGTGA 1440  
DB 1381 ACAGTGAAGTGTATGAAGCAATTTAGTGCATCAAAACAGATGAAGATGAAGTGA 1440  
QY 1441 CGCCCAATATATTTAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1500  
DB 1441 CGCCCAATATATTTAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1500  
QY 1501 AAATGACAGTATGATGAATTAACGACAAAGATGTTATGCTGAGGCGCGAGGAGC 1560  
DB 1501 AAATGACAGTATGATGAATTAACGACAAAGATGTTATGCTGAGGCGCGAGGAGC 1560  
QY 1561 TCTTAAATGAGCCAC 1620  
DB 1561 TCTTAAATGAGCCAC 1620  
QY 1621 GACCCCTGCTGATCAGAC 1680  
DB 1621 GACCCCTGCTGATCAGAC 1680  
QY 1681 AATGCTGATCAGAC 1740  
DB 1681 AATGCTGATCAGAC 1740  
QY 1741 CTGCGGAGCAAGCTTGGGTTAATTAAGAGGCTGACACATCTGAGGAGGAAATAGT 1800  
DB 1741 CTGCGGAGCAAGCTTGGGTTAATTAAGAGGCTGACACATCTGAGGAGGAAATAGT 1800  
QY 1801 TCCGAGACAGACAAATTCACATAGCGATTTCTCCCAATTAAGAAATGATCAGTGC 1860  
DB 1801 TCCGAGACAGACAAATTCACATAGCGATTTCTCCCAATTAAGAAATGATCAGTGC 1860  
QY 1861 GATCTTACCTTTTCCCAACCTTTTAAATTAAGAGGCTTGAACACCTTTAGAA 1920  
DB 1861 GATCTTACCTTTTCCCAACCTTTTAAATTAAGAGGCTTGAACACCTTTAGAA 1920  
QY 1921 GAAAAAAGAAATCCCTTGGCCCGGCGCAATTCACATTCGGAAGAGCTACCAATTA 1980  
DB 1921 GAAAAAAGAAATCCCTTGGCCCGGCGCAATTCACATTCGGAAGAGCTACCAATTA 1980

QY 1981 AGAATTTTCCCTTTTATGAGGGGGGAGCTCTTTTCTTATCTTGGGGTGGGATTTG 2040  
D 1981 AGAATTTTCCCTTTTATGAGGGGGGAGCTCTTTTCTTATCTTGGGGTGGGATTTG 2040  
QY 2041 GCGCCCAACGAGGAGCTTTTATTTATTTATGATGATCTGGGATTTTCCCTGGCTA 2100  
D 2041 GCGCCCAACGAGGAGCTTTTATTTATTTATGATGATCTGGGATTTTCCCTGGCTA 2100  
QY 2101 CTATTTATCCAGCGCGCCAGCCCAAAATATTTAAACAGAGATCTTAAGATGGCGGT 2160  
D 2101 CTATTTATCCAGCGCGCCAGCCCAAAATATTTAAACAGAGATCTTAAGATGGCGGT 2160  
QY 2161 ACATTTACCAACACCTGCTGGAGGAGGATGATATCCCTCAATACCACTTTCTGTTC 2220  
D 2161 ACATTTACCAACACCTGCTGGAGGAGGATGATATCCCTCAATACCACTTTCTGTTC 2220  
QY 2221 CCGCAATGGAATAATGATCCGAGAGGCCATCTTTTGGGGGAGATTCATTTACTA 2280  
D 2221 CCGCAATGGAATAATGATCCGAGAGGCCATCTTTTGGGGGAGATTCATTTACTA 2280  
QY 2281 TTACCTTTACCAAAAGACTATCCAGATGGAATGTAACGATGGCGCATGCAACGTC 2340  
D 2281 TTACCTTTACCAAAAGACTATCCAGATGGAATGTAACGATGGCGCATGCAACGTC 2340  
QY 2341 CGAAGATTTTATGATGAGAGGAGGAGGATGATCCGATCCGAGATATACCAATTA 2400  
D 2341 CGAAGATTTTATGATGAGAGGAGGAGGATGATCCGATCCGAGATATACCAATTA 2400  
QY 2401 TGGTGATC 2408  
D 2401 TGGTGATC 2408

## RESULT 2

US-09-317-179-5  
Sequence 5, Application US/09317179  
Patent No. 6383769

GENERAL INFORMATION:

APPLICANT: Tsusaki et al.  
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA

ZIP: 20004  
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/317,179  
FILING DATE: 24-May-1999  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/870,827

FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TSUSAKI-2  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2408 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double

STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Bacillus sp.  
INDIVIDUAL ISOLATE: V230 (FERM BP-5054)  
NAME/KEY: signal peptide  
LOCATION: 1..360  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-317-179-5

Query Match 100.0%; Score 2408; DB 4; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGGAATAATCTAGATTTCAATTTGGCCAGACTTCCAGTTGGTGTAAAGAGAGTTGG 60  
D 1 CCGGGAATAATCTAGATTTCAATTTGGCCAGACTTCCAGTTGGTGTAAAGAGAGTTGG 60  
QY 61 GACTGCCAATGAGCTGTGGGTAAGAAACAGCTTACTCATGAGCAATTACTAGAGAGAT 120  
D 61 GACTGCCAATGAGCTGTGGGTAAGAAACAGCTTACTCATGAGCAATTACTAGAGAGAT 120  
QY 121 TTCAAAAGCTCTGGGATTAAGCCAAAGTCCACTTTGAAATTAACCTTTTCAGCCTGTGTG 180  
D 121 TTCAAAAGCTCTGGGATTAAGCCAAAGTCCACTTTGAAATTAACCTTTTCAGCCTGTGTG 180  
QY 181 GGGGCTTTTGTGTTTATTTATTTTCAACTGCAGAGTGTCCATCCCTATATCAATTTAA 240  
D 181 GGGGCTTTTGTGTTTATTTATTTTCAACTGCAGAGTGTCCATCCCTATATCAATTTAA 240  
QY 241 GACGAATTTTATTCATTCATCCATCCATCCCAATTAATGCTGCTCTATATCTTTAA 300  
D 241 GACGAATTTTATTCATTCATCCATCCATCCCAATTAATGCTGCTCTATATCTTTAA 300  
QY 301 TTAATAAGAACTATCAAGAGCTTTCTTATCAAAATTCATATATCAAGAGAGGAGAGC 360  
D 301 TTAATAAGAACTATCAAGAGCTTTCTTATCAAAATTCATATATCAAGAGAGGAGAGC 360  
QY 361 ATGAATCTCAAAAGATTTGGCGAAACAGAGTGGCGTAACTTTAGAGCTGTATATTA 420  
D 361 ATGAATCTCAAAAGATTTGGCGAAACAGAGTGGCGTAACTTTAGAGCTGTATATTA 420  
QY 421 GTAGAGCGGAGCGGACGCATATTTTGGCGAGCAATGAACAGGCGGAGCTACAGAGAA 480  
D 421 GTAGAGCGGAGCGGACGCATATTTTGGCGAGCAATGAACAGGCGGAGCTACAGAGAA 480  
QY 481 GACTATGCTTTTGGCCATATTACACGCGCTGACATGCTAATAAATTCAGAGCAACAAAC 540  
D 481 GACTATGCTTTTGGCCATATTACACGCGCTGACATGCTAATAAATTCAGAGCAACAAAC 540  
QY 541 AGTCTCAATTTAAAGTGCCTCAATTCATTAATGATAGCAATCAAAACATTGATTGGCA 600  
D 541 AGTCTCAATTTAAAGTGCCTCAATTCATTAATGATAGCAATCAAAACATTGATTGGCA 600  
QY 601 AAAGGCTATGATTAAGTCAAGCACTTATAGATTAGATGATGAGAGTGGCGACATG 660  
D 601 AAAGGCTATGATTAAGTCAAGCACTTATAGATTAGATGATGAGAGTGGCGACATG 660  
QY 661 CAAAACGCTGATGCTACTGCGCAAAATTTATGATGATATCAATGCTCCGCTTTAGCA 720  
D 661 CAAAACGCTGATGCTACTGCGCAAAATTTATGATGATATCAATGCTCCGCTTTAGCA 720  
QY 721 GGTGACCCCAAAACAGATGATATCTCACTTCTATTTATCTATCAAAAGTGGCTGAT 780  
D 721 GGTGACCCCAAAACAGATGATATCTCACTTCTATTTATCTATCAAAAGTGGCTGAT 780  
QY 781 ACATGATTTGACAGCTGAGAAATGCTGAGAGATTTGAAGATGATGATGATGATTTGTT 840  
D 781 ACATGATTTGACAGCTGAGAAATGCTGAGAGATTTGAAGATGATGATGATGATTTGTT 840  
QY 841 CCAATGATCGGTATCTTAATATCAACACAGAGAGTGTGCTGCTACTTTAACC 900

```

Db      841  CCAATGATCCGCTATCTTAATATCAAAACACAGAGTGGTCTGCTCTACTTAAC  900
Qy      901  AAAGATGGCCAGACCCGTTTATCTATCAGATTCCTAGTAATCCGAATGCTGGA  960
Db      901  AAAGATGGCCAGACCCGTTTATCTATCAGATTCCTAGTAATCCGAATGCTGGA  960
Qy      961  ACCGCTGCTGTATCAATCATTCCTCACTGCTCAAGTAATCTTCCAGCGGATCA  1020
Db      961  ACCGCTGCTGTATCAATCATTCCTCACTGCTCAAGTAATCTTCCAGCGGATCA  1020
Qy      1021  GCTACACTTAAAGTGTAGTATCTGATCATTAATCTGCTTGTATGCGCGAAGCGT  1080
Db      1021  GCTACACTTAAAGTGTAGTATCTGATCATTAATCTGCTTGTATGCGCGAAGCGT  1080
Qy      1081  ACAGTTTATCAAAATATTCAGCAATTTATGATGAGGCAAGTGAATTCAGGTATAC  1140
Db      1081  ACAGTTTATCAAAATATTCAGCAATTTATGATGAGGCAAGTGAATTCAGGTATAC  1140
Qy      1141  CATACTTTAAGAGACCCCTCACTATGTGAGATAGAGGCAATTAATCTTCTTGA  1200
Db      1141  CATACTTTAAGAGACCCCTCACTATGTGAGATAGAGGCAATTAATCTTCTTGA  1200
Qy      1201  GCGAATCTGGAACAACAGATGTTATCAAGCGCATCAGCTTCAATTAATAGCTTAC  1260
Db      1201  GCGAATCTGGAACAACAGATGTTATCAAGCGCATCAGCTTCAATTAATAGCTTAC  1260
Qy      1261  TATGCGGAGTGAAGCTCTTCTTCCAGATGAAAAAATTAATCTGCTTCAAAAGTCTAA  1320
Db      1261  TATGCGGAGTGAAGCTCTTCTTCCAGATGAAAAAATTAATCTGCTTCAAAAGTCTAA  1320
Qy      1321  AAACAATTTGCTTCTTATAGCGAATGTCATTAGGCAATTTGAAATGGCCGATGACTAT  1380
Db      1321  AAACAATTTGCTTCTTATAGCGAATGTCATTAGGCAATTTGAAATGGCCGATGACTAT  1380
Qy      1381  ACAGTGAAGAGTGTATGAACCATTAAGTCCGATCAAAACAGTATGAGATTAATCGAA  1440
Db      1381  ACAGTGAAGAGTGTATGAACCATTAAGTCCGATCAAAACAGTATGAGATTAATCGAA  1440
Qy      1441  CGCGCCATATATTAATGAATGAATGAATGATTAATGATTAATGATTAATGATTAATG  1500
Db      1441  CGCGCCATATATTAATGAATGAATGAATGATTAATGATTAATGATTAATGATTAATG  1500
Qy      1501  AAAATGACGAGTGTATGAATTAAGCAAAAGATGTTATATCTAGGCGCGGAGCGAC  1560
Db      1501  AAAATGACGAGTGTATGAATTAAGCAAAAGATGTTATATCTAGGCGCGGAGCGAC  1560
Qy      1561  TCCTTAATGAGCCGACACACACCCGATTAATGAATGAACTGCACTTGTATGAACTGATCTT  1620
Db      1561  TCCTTAATGAGCCGACACACACCCGATTAATGAATGAACTGCACTTGTATGAACTGATCTT  1620
Qy      1621  GACCTGCTGATCTACACACTTACTCTCATTCGCGATCCGACCCGACCTGAAGTAT  1680
Db      1621  GACCTGCTGATCTACACACTTACTCTCATTCGCGATCCGACCCGACCTGAAGTAT  1680
Qy      1681  AATGTGTACTACACAGTTATATGAGCAATAGGCTTCTATCGAATCATCTCTAC  1740
Db      1681  AATGTGTACTACACAGTTATATGAGCAATAGGCTTCTATCGAATCATCTCTAC  1740
Qy      1741  CTGGGGGACAGCTGGGGTTAATTAAGGCTGTGACACATCTGGAGGAAATATGCT  1800
Db      1741  CTGGGGGACAGCTGGGGTTAATTAAGGCTGTGACACATCTGGAGGAAATATGCT  1800
Qy      1801  TCCGACAGAGCAATTCCTATAGGATTAATTCCTCAATTAAGAAATGCTACTGCA  1860
Db      1801  TCCGACAGAGCAATTCCTATAGGATTAATTCCTCAATTAAGAAATGCTACTGCA  1860
Qy      1861  GATCTTACTTTTCCCAACTTTTAAATTAAGGCTTGTATGACACCTTTAGAA  1920
Db      1861  GATCTTACTTTTCCCAACTTTTAAATTAAGGCTTGTATGACACCTTTAGAA  1920
Qy      1921  GAAAAAATTCCTTGGCCGCAATTTACCATTTGCCGAGGAGCTTACCAATTAATA  1980
Db      1921  GAAAAAATTCCTTGGCCGCAATTTACCATTTGCCGAGGAGCTTACCAATTAATA  1980

```

```

Db      1921  GAAAAAATTCCTTGGCCGCGGCAATTTACCATTTGCCGAGGAGCTTACCAATTAATA  1980
Qy      1981  AGAAATTTCCCTTTTATGAGGAGGAGCTCTTTTCTATCTTGGGAGGATGCTG  2040
Db      1981  AGAAATTTCCCTTTTATGAGGAGGAGCTCTTTTCTATCTTGGGAGGATGCTG  2040
Qy      2041  GCGCCACAGGAGCTTTTATTAATTAATGATGATTCGCGATTTTCCGCGCTA  2100
Db      2041  GCGCCACAGGAGCTTTTATTAATTAATGATGATTCGCGATTTTCCGCGCTA  2100
Qy      2101  CTTATTTACCAAGCCCGCCAGCAAAATTAATTAACAGAGATCTTAAGTAGCGCT  2160
Db      2101  CTTATTTACCAAGCCCGCCAGCAAAATTAATTAACAGAGATCTTAAGTAGCGCT  2160
Qy      2161  ACAGTTTACCAACCTCTGAGGAGGATGATTCGCGATTCGCAATTCATCTTCTGCTC  2220
Db      2161  ACAGTTTACCAACCTCTGAGGAGGATGATTCGCGATTCGCAATTCATCTTCTGCTC  2220
Qy      2221  CCGACAAATGCAAAATGATCCGAGAGGCCATCTTTTGGGGAGATATCATTTACTA  2280
Db      2221  CCGACAAATGCAAAATGATCCGAGAGGCCATCTTTTGGGGAGATATCATTTACTA  2280
Qy      2281  TTACCTTTACCAACAAGCTATCCAGATGGAATGTCAGGAATGCGCGCATGCACTC  2340
Db      2281  TTACCTTTACCAACAAGCTATCCAGATGGAATGTCAGGAATGCGCGCATGCACTC  2340
Qy      2341  CGAAGATTTATGCTTGTGAGAGGAGGAGGATTCGCTCCGAATTTACCAATTAATA  2400
Db      2341  CGAAGATTTATGCTTGTGAGAGGAGGAGGATTCGCTCCGAATTTACCAATTAATA  2400
Qy      2401  TGTGATC 2408
Db      2401  TGTGATC 2408

```

```

RESULT 3
US-08-870-827-4
; Sequence 4, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Teusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1365 base pairs

```

TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: Linear  
US-08-870-827-4

Query Match 56.7%; Score 1365; DB 2; Length 1365;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

457 ATGAACGGGGGACTACAGAGACATATGTTTGGCCATATATACACGGCGTACATG 516  
|||||  
1 ATGAACGGGGGACTACAGAGAGACTATGTTTGGCCATATATACACGGCGTACATG 60  
517 CTAATAATCCAGAGACAACAACAGTCTCAATTTAAAGTGCCCAATTAATGCATCA 576  
61 CTAATAATCCAGAGACAACAACAGTCTCAATTTAAAGTGCCCAATTAATGCATCA 120  
577 GCAATCAAAAACATTGATTCGGCAAAAGGTAATAGTACAGCCACTTAATGATTTA 636  
|||||  
121 GCAATCAAAAACATTGATTCGGCAAAAGGTAATAGTACAGCCACTTAATGATTTA 180  
637 GATGATGGATAGTACCTGGCCACTGCAAAAGCTGATGTACTGGCGCAATTAATGGA 696  
|||||  
181 GATGATGGATAGTACCTGGCCACTGCAAAAGCTGATGTACTGGCGCAATTAATGGA 240  
697 TATCAGATCTCTCCGCTTTAGCAGGTGACCCAAAACAGTATGATCTCCACTTCAT 756  
|||||  
241 TATCAGATCTCTCCGCTTTAGCAGGTGACCCAAAACAGTATGATCTCCACTTCAT 300  
757 TTATCTATCAAAAAGTCGTGATATCATGATGACAGCTGGAAAAATGCTGGAAGATA 816  
|||||  
301 TTATCTATCAAAAAGTCGTGATATCATGATGACAGCTGGAAAAATGCTGGAAGATA 360  
817 TTGAAGATATGATTAATTTGTTCCAAATGATCCGTATCTTAATATCAACACAGAG 876  
|||||  
361 TTGAAGATATGATTAATTTGTTCCAAATGATCCGTATCTTAATATCAACACAGAG 420  
877 TGTGAGGTCTGCTACTTTAAACCAAGATGGCCAGTCCGTTATTTCTATACAGATAC 936  
|||||  
421 TGTGAGGTCTGCTACTTTAAACCAAGATGGCCAGTCCGTTATTTCTATACAGATAC 480  
937 TCAGATATCTCTGAAGATGTTGGAACCGGCTCTGTTAACCAAAATCAATTCATGCTCAA 996  
|||||  
481 TCAGATATCTCTGAAGATGTTGGAACCGGCTCTGTTAACCAAAATCAATTCATGCTCAA 540  
997 GTAACTATATCCACGGCGGATGACGCTACTTAAGTCGATGGAGTCTGATCATATAA 1056  
|||||  
541 GTAACTATATCCACGGCGGATGACGCTACTTAAGTCGATGGAGTCTGATCATATAA 600  
1057 TCTGCTTTGATGGCGAGAGCGTACAGTTATCAAAATATTCAGCAATTTATCGATGAA 1116  
|||||  
601 TCTGCTTTGATGGCGAGAGCGTACAGTTATCAAAATATTCAGCAATTTATCGATGAA 660  
1117 GCGAAGTGAGATTTCAGGTGATACCATATCTTTAAGAGACCTCACTATGTTGAAGATAAG 1176  
|||||  
661 GCGAAGTGAGATTTCAGGTGATACCATATCTTTAAGAGACCTCACTATGTTGAAGATAAG 720  
1177 GCGCATTAATATCTGCTTTGAGCGCAATACCTGGAACAACAGATGGTTATCAAGGCAT 1236  
|||||  
721 GCGCATTAATATCTGCTTTGAGCGCAATACCTGGAACAACAGATGGTTATCAAGGCAT 780  
1237 CAGTCTTTCAATTAAGTCTTACTATGCGGAAGTACGCTCTTCTTCCAGATGAATAA 1296  
781 CAGTCTTTCAATTAAGTCTTACTATGCGGAAGTACGCTCTTCTTCCAGATGAATAA 840  
1297 AATAAAGCTCTCAAAAGTCTTAACAAATGCTTCTTACGAATGGTCAATAGGC 1356  
841 AATAAAGCTCTCAAAAGTCTTAACAAATGCTTCTTACGAATGGTCAATAGGC 900  
1357 ATTGTTGAATGGCGGATGATATACAGTGAAGTGTATGAACCAATTAAGTGCATCA 1416  
901 ATTGTTGAATGGCGGATGATATACAGTGAAGTGTATGAACCAATTAAGTGCATCA 960

1417 AACACAGTAGAGATGAAGTCGAACGCCCATATATTTAAATGAATAAATGAT 1476  
961 AACACAGTAGAGATGAAGTCGAACGCCCATATATTTAAATGAATAAATGAT 1020  
1477 CTATTCACGAGATTCAAGAGATTCACAAATGACAGAGTATGGAATTAACGACAAAGATGT 1536  
1021 CTATTCACGAGATTCAAGAGATTCACAAATGACAGAGTATGGAATTAACGACAAAGATGT 1080  
1537 TATATGCTAAGGCGCGGAGGCGACCTCTTAATATGCGCCACACACCCGATTAATGAACCT 1596  
1081 TATATGCTAAGGCGCGGAGGCGACCTCTTAATATGCGCCACACACCCGATTAATGAACCT 1140  
1597 GCACTTGTATGAACATCAATCTTGACCCCTGCTGATCTCACACACACTTACTCTATTGC 1656  
1141 GCACTTGTATGAACATCAATCTTGACCCCTGCTGATCTCACACACACTTACTCTATTGC 1200  
1657 GGTATCCGACCCCTGGAAGTATATATGCTGCTACACAAAGTTATATGCAATAGAGGC 1716  
1201 GGTATCCGACCCCTGGAAGTATATATGCTGCTACACAAAGTTATATGCAATAGAGGC 1260  
1717 TTCTATCCGAACATCACTCTACCTGCGGCAAGCTTGGGTTAATTAAGGCTCT 1776  
1281 TTCTATCCGAACATCACTCTACCTGCGGCAAGCTTGGGTTAATTAAGGCTCT 1320  
1777 GACACATCTGAGAGAAATAGTCCGAGACAGACAAATTCACA 1821  
1321 GACACATCTGAGAGAAATAGTCCGAGACAGACAAATTCACA 1365

RESULT 4  
US-09-317-179-4  
Sequence 4, Application us/09317179  
Patent No. 6383769  
GENERAL INFORMATION:  
Applicant: Tsusaki et al.  
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEWMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/317,179  
FILING DATE: 24-May-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/870,827  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TSUSAKI=2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1365 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-317-179-4

Query Match 56.7%; Score 1365; DB 4; Length 1365;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 457 ATGACACGGGGGACTACAGAGAGACTAGTTTGGCCATATTACAGCGCGTACATG 516
DB 1 ATGAACACGGGGGACTACAGAGAGACTAGTTTGGCCATATTACAGCGCGTACATG 60
QY 517 CTAATAATTCAGGCAACAACAGTCTCAATTAAGTGCCCAATTCATGATCA 576
DB 61 CTAATAATTCAGGCAACAACAGTCTCAATTAAGTGCCCAATTCATGATCA 120
QY 577 GCAATCAAAAACATTGATTGCGCAAAAGGATGATGATGATGATGATGATGATGAT 636
DB 121 GCAATCAAAAACATTGATTGCGCAAAAGGATGATGATGATGATGATGATGATGAT 180
QY 637 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 696
DB 181 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 697 TATCATCATGCTCCGCTTTAGCAGGTGACCCCAAAAACAGTATGATGATGATGAT 756
DB 241 TATCATCATGCTCCGCTTTAGCAGGTGACCCCAAAAACAGTATGATGATGATGAT 300
QY 757 TTATTTCTATCAAAAAGTGGTGATACATGATGATGATGATGATGATGATGATGAT 816
DB 301 TTATTTCTATCAAAAAGTGGTGATACATGATGATGATGATGATGATGATGATGAT 360
QY 817 TTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876
DB 361 TTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 877 TGTGAGGTGCTGCTACTTTTAAACAAGATGCGCAAGTCCGTTTATTCATGATGAT 936
DB 421 TGTGAGGTGCTGCTACTTTTAAACAAGATGCGCAAGTCCGTTTATTCATGATGAT 480
QY 937 TCAGTAAATCTCGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 996
DB 481 TCAGTAAATCTCGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 997 GTAACCTATCCAGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1056
DB 541 GTAACCTATCCAGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 1057 TCTGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1116
DB 601 TCTGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 1117 GGCAGTGGATTTAGGTGATTAACCAATCTTTAAGAGACCTGCTATGTTGAAGTA 1176
DB 661 GGCAGTGGATTTAGGTGATTAACCAATCTTTAAGAGACCTGCTATGTTGAAGTA 720
QY 1177 GGCATTAATATCTGCTTTGAAGCGAATGATGATGATGATGATGATGATGATGAT 1236
DB 721 GGCATTAATATCTGCTTTGAAGCGAATGATGATGATGATGATGATGATGATGAT 780
QY 1237 CAGTCTTTCAATATAAGCTTACTTATGCGGAGTGAAGTCTTCTTCCAGATGA 1296
DB 781 CAGTCTTTCAATATAAGCTTACTTATGCGGAGTGAAGTCTTCTTCCAGATGA 840
QY 1297 AATAAAGCTTCAAAAGTCTTAAACAAATGCTTCTTTAAGCAATGCTGATAGGC 1356
DB 841 AATAAAGCTTCAAAAGTCTTAAACAAATGCTTCTTTAAGCAATGCTGATAGGC 900
QY 1357 ATTGTTGATGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1416
DB 901 ATTGTTGATGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 1417 AACACAGTAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1476
DB 961 AACACAGTAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1477 CTATTCACGATTTCAAGAGATCAAAATGACGAGTATGATGATTAACGACAAAGAT 1536
  
```

```

DB 1021 CTATTCACGATTTCAAGAGATCAAAATGACGAGTATGATGATTAACGACAAAGATGTT 1080
QY 1537 TATATGCTAGGGCCCGGAGGGGACTCCTTAATATGCGCCACACACCGGATTAATGA 1596
DB 1081 TATATGCTAGGGCCCGGAGGGGACTCCTTAATATGCGCCACACACCGGATTAATGA 1140
QY 1597 GGACTTGTATTAACATGAATCTTGACCTCTGATCTGATCTGATCTGATCTGATCTGAT 1656
DB 1141 GGACTTGTATTAACATGAATCTTGACCTCTGATCTGATCTGATCTGATCTGATCTGAT 1200
QY 1657 GGTATCCGCAACCTGGAAGTAAATGATGATGATGATGATGATGATGATGATGAT 1716
DB 1201 GGTATCCGCAACCTGGAAGTAAATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1717 TTTATTCAGAAACATCATCTGCTGCTGCGGACAAAGTGGGGTTAATTAAGGGTCT 1776
DB 1261 TTTATTCAGAAACATCATCTGCTGCTGCGGACAAAGTGGGGTTAATTAAGGGTCT 1320
QY 1777 GACACATCTGGAGGAGAAATAGTTCCGACAGACAAATTC 1821
DB 1321 GACACATCTGGAGGAGAAATAGTTCCGACAGACAAATTC 1365
  
```

```

RESULT 5
US-09-058-746-1/c
: Sequence 1, Application, US/09058746
: Patent No. 6022716
:
GENERAL INFORMATION:
APPLICANT: Ilya Chumakov
APPLICANT: Hiroaki Tanaka
TITLE OF INVENTION: High Throughput DNA sequencing vector
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/058,746
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: GENSET.015A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10317 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
MOLECULE TYPE: synthetic DNA
ORIGINAL SOURCE:
ORGANISM: Cloning vector pGendel
FEATURE:
NAME/KEY: pGendel1
LOCATION: 1..10317
FEATURE:
NAME/KEY: Homology with X06404 comp1 (411..1668)
LOCATION: 9..1266
IDENTIFICATION METHOD: blastn against X06404
FEATURE:
  
```



```
NAME/KEY: Kanamycin resistance gene CDS
LOCATION: 142..957
IDENTIFICATION METHOD: By homology to X06404
FEATURE:
NAME/KEY: Tn1000's right end
LOCATION: complement 1332..1371
IDENTIFICATION METHOD: blastn against X60200)
FEATURE:
NAME/KEY: Homology with U46017 (1-472)
LOCATION: 1423..1894
IDENTIFICATION METHOD: blastn against U46017
FEATURE:
NAME/KEY: single stranded DNA replication origin
LOCATION: 1423..1894
IDENTIFICATION METHOD: By homology to U46017
OTHER INFORMATION: mutation T -> C 1658
FEATURE:
NAME/KEY: Homology with U51113 (2382..6997)
LOCATION: 1896..6544
IDENTIFICATION METHOD: blastn against U51113
FEATURE:
NAME/KEY: OriS
LOCATION: 1972..2188
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: repELR
LOCATION: 2897..2918
OTHER INFORMATION: Described in seqid 16
FEATURE:
NAME/KEY: RepE
LOCATION: 2903..3034
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: T3
LOCATION: 3043..3059
OTHER INFORMATION: Described in seqid 17
FEATURE:
NAME/KEY: LRT3RA
LOCATION: complement 3045..3069
OTHER INFORMATION: Described in seqid 15
FEATURE:
NAME/KEY: IncC
LOCATION: 3070..3320
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: Insertion 33 bases 3038..3071
FEATURE:
NAME/KEY: Para
LOCATION: 3655..4821
IDENTIFICATION METHOD: By homology to U51113
OTHER INFORMATION: mutation G -> A 3878
FEATURE:
NAME/KEY: ParB
LOCATION: 4821..5792
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: ParC
LOCATION: 5865..6382
IDENTIFICATION METHOD: By homology to U51113
FEATURE:
NAME/KEY: Homology with J01688 (complement 175..819)
LOCATION: 6574..7218
IDENTIFICATION METHOD: blastn against J01688
OTHER INFORMATION: mutation A -> G 7096
FEATURE:
NAME/KEY: CDS streptomycin sensitivity gene
LOCATION: complement 6716..7090
IDENTIFICATION METHOD: By homology to J01688
OTHER INFORMATION: mutation A -> G 6728
OTHER INFORMATION: mutation G -> C 6821
OTHER INFORMATION: mutation C -> T 6866
OTHER INFORMATION: mutation T -> C 7013
OTHER INFORMATION: mutation T -> A 7058
FEATURE:

NAME/KEY: ipSLR
LOCATION: 7155..7174
OTHER INFORMATION: Described in seqid 12
FEATURE:
NAME/KEY: SP6
LOCATION: 7230..7248
OTHER INFORMATION: Described in seqid 13
FEATURE:
NAME/KEY: Tn1000's left end
LOCATION: 7252..7291
IDENTIFICATION METHOD: blast (X60200)
FEATURE:
NAME/KEY: Homology with X02730 (complement 37..1959)
LOCATION: 7305..9227
IDENTIFICATION METHOD: blastn against X02730
FEATURE:
NAME/KEY: CDS levansucrase gene
LOCATION: complement 7379..8800
IDENTIFICATION METHOD: By homology to X02730
OTHER INFORMATION: mutation T -> C 7466
OTHER INFORMATION: mutation A -> G 7739
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347
OTHER INFORMATION: mutation T -> C 8600
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772
FEATURE:
NAME/KEY: SLR3
LOCATION: 8711..8731
OTHER INFORMATION: Described in seqid 14
FEATURE:
NAME/KEY: Homology with J01636 (complement 1158..1465)
LOCATION: 9298..9623
IDENTIFICATION METHOD: blastn against J01636
FEATURE:
NAME/KEY: CDS alpha peptide beta-galactosidase
LOCATION: complement 9276..9497
IDENTIFICATION METHOD: By homology to J01636
FEATURE:
NAME/KEY: primer HE1
LOCATION: complement 9465..9479
FEATURE:
NAME/KEY: primer HE2
LOCATION: 9461..9475
FEATURE:
NAME/KEY: primer lacRS2AVT
LOCATION: complement 9603..9630
FEATURE:
NAME/KEY: primer lacE2M1u
LOCATION: 9289..9314
FEATURE:
NAME/KEY: Homology with M77789 (1889..2576)
LOCATION: 9629..10315
IDENTIFICATION METHOD: blastn against M77789
FEATURE:
NAME/KEY: high copy-number double-stranded DNA replication origin
LOCATION: complement 9629..10315
IDENTIFICATION METHOD: By homology to M77789
OTHER INFORMATION: mutation C -> T 9803
OTHER INFORMATION: site ScaI 10029 - 10034
OTHER INFORMATION: site PmlI 10038 - 10043
OTHER INFORMATION: CLONING SITES 10031 - 10041
FEATURE:
NAME/KEY: oriLRd
LOCATION: 9856..9881
OTHER INFORMATION: Described in seqid 8
FEATURE:
NAME/KEY: OSI
LOCATION: 10009..10026
OTHER INFORMATION: Described in seqid 10
FEATURE:
NAME/KEY: ORI
LOCATION: complement 10046..10062
OTHER INFORMATION: Described in seqid 11
FEATURE:
```

NAME/KEY: or1.rn1  
 LOCATION: complement 10182..10202  
 OTHER INFORMATION: Described in seqid 9  
 US-09-058-746-1

Query Match 25.2%; Score 607.2; DB 3; Length 10317;  
 Best Local Similarity 66.3%; Pred. No. 1.5e-158;  
 Matches 993; Conservative 0; Mismatches 448; Indels 57; Gaps 6;

```

QY 358 ACGAGACCTCAAAAGATGGCGGAAAAGACAGTGGCCCTAACCTTCAGACCTCTCTA 417
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8803 ACGAGAACATCAAAAGATTTGGCAAAACAGTAACAGATTAACCTTACTACCGCAGC 8744
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 418 TTAGTAGGAGCGGAGACCGCATATTTTGGCGAGCAATGAACGGGGAGCTACAG 477
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8743 CTGGAGGAGAGG---CGCACTCAACGGTTTGGGAAGAAGAACACAAAGCATATAG 8687
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 478 GAAGACTATGTTTGGCCATATTAACGCGCTGACATCTAAATTCAGAGACAA 537
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8686 GAACATACGGCATTTCCCATATTAACGCCATGATGTCGCAATCCCTGAAACGCA 8627
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 538 AACAGTCCCATTTAAAGTGCCCTCATTCATGATGACGAATCAAAACATTGATCG 597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8626 AAAATGAAAAATATCAAGTTCTTCAAGTTGCTCACAATTAATAATATCTCTCT 8567
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 598 GCAAAAGGCTATGATAGTCAAGGCACCTTAATAGATTAGATTGGGATAGCTGGCA 657
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8566 GCAAAAGG-----CTTGGAGCTTTGGGAGACGCTGGGCA 8534
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 658 CTGCAAAACGCTGATGCTACTGCGCAATTAATCTGATATCAGATGCTCTCCCTTTA 717
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8533 TTACAAAAAGCTGAGCGACGCTGCGCAACATACGCGCTACCCATGCTCTTTCATTA 8474
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 718 GCAGGTGACCCCAAAACAGTATGATGATCTCACTTCTTATCTATCAAAAGTGGT 777
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8473 GCGGAGATTCCTAAAAATGCGGATGACATGATTTTCAATGTTTATCAAAAGTGGC 8414
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 778 GATACATGATTTGACGCTGAAAAAATGCTGAGAGATTTTGAAGATATGATTAATTT 837
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8413 GAAACTTTATTTGACGCTGAAAAAAGCTGCGCCCTTTTAAAGACAGGACAAATTC 8354
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 838 GTTCCAATGATCCGCTATCTTAATATCAACACAGAGTGGTCTGCTCTACTTTA 897
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8353 GATGCAATGATTTCTATCTTAAGACCAACACAGAGTGGTCTGCTGATCCCATTT 8294
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 898 ACCAAGATGGCAAGTCCGTTTATTTATACAGATTACTAGTAATCTGAAGATGCT 957
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8293 ACATCTGACGAAAAATCCGTTTATTTCTACACTGATTTCTCGGTAAACATTA----- 8241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 958 GGAACGGGTGCTGTAACCAATCTTTCACTGCTCAAGTAACCTTATCCAGCGGAT 1017
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8240 -----CGGCAAAACAACTGACAACTGACAAAGTTAAGCTATCAGCATCAGAC 8192
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1018 GCAGCTACCTTAAAGTGAAGTGAATCTGATCTAATATCTGTTTGAAGCGGAGAC 1077
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8191 AGCTCT---TTGAATCATACGGTGTAGAGATTATTAATCAATTTTTGA---CGGTAC 8138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1078 GGTACAGTTTATCAAAATATTCAGCAATTTATCGATGAAGCAAGTGAATTCAGTGA 1137
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8137 GGAANAAGCTATCAAAATGCTACAGAGTTCAATGATGAAGCACTAACGCTCAGCGAC 8078
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1138 AACCATCTTTAAGACCTCTACTATGTTGAAGTAAAGGCCATTAATCTGCTCTT 1197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8077 AACCATCTGAGAGATCTCTACACTAGTAGAAGTAAAGGCCAACAAATCTACTATTT 8018
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1198 GAAGGATATCTGGAACAACAGATGTTATCAAGCGCATCTTTCAATATTAAGCT 1257
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8017 GAAGCAAACTGGAAGTGAAGATGGCTACCAAGGCAAGATCTTTTAAACAAACA 7958
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1258 TACTATGGCGAAGTGAAGTCTTCTCCAGATGAAGAAAAATTAACCTGCTTCAAGTCT 1317
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7957 TACTATGGCAAAAGCACATCTTTCCGCTCAAGAAAGTCAAAACCTTGCAGAGCAT 7898
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 1318 AAAAAACAATTCCTCTTACGCAATGCTGATTAAGCATTTGTAATGGCCGATGAC 1377
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7897 AAAAAACCGCTGATTAAGCAAGCGGCTCTCGTATGATTAAGCTAAAGATGAT 7838
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1378 TATACAGTGAAGTGTATGAACCATTAAGTCTGATCAACAGATTAAGATGATC 1437
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7837 TACACACTGAAGAAAGTGAAGTAAACCGCTGATTCATCTAACAGATGAAGATTT 7778
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1438 GAACGGCCCATATTAATTAATGAATTAATGAATGATTAATGATTAAGATTAAGGA 1497
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7777 GAACGGCGAACGCTTTAAATGAACGCGCAAAATGATGATCTGCTACACTGCTCCGCGA 7718
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1498 TCCAAAATGACAGTGAATTAACGACAAAGATGTTATATGATAGGCGCGGAGGC 1557
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7717 TCAAAATGACAGATTAAGCAGCATTAAGCTTAAGCTATTAATGATGATGATTTCT 7658
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1558 GACTCTTAATGCGCCACACACCCGTAATTAATGAAGTGAAGTGAATTAATGAT 1617
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7657 AATCTTTAAGTGGCCCTACACAGCCGCTGAACAAACTGCGCTTGTGTTAAATGAT 7598
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1618 CTGACCCCTGCTGATCTCACACACTTACTCTGATTTGGGATCCGACCCGTAAGT 1677
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7597 CTGATCTTAAGATTAAGTAACTTACTTACTTACTTACTTACTTACTTACTTACTT 7538
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1678 AATAATGTGACTCACAGTTATATGACAGATAGAGGCTTCTATCCAGAACATCACT 1737
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7537 AACATGCTGATTAAGCTATATGACAAACAGAGATTTCTACGCAACAAATCA 7478
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1738 CACCTGGCGGACAACTTGGGTTAATTAATGAAGGCTGACATCTGAGGAGAAAT 1797
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7477 ACGTTTGGCGGACCTTCTCTGACATCAAAAGGCAAGAAACATCTGTTGCAAGAC 7418
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1798 AGTTCC---GGACAGACCAATTCCTCAAGGATTAATCTCCCAATAAAGAAATGTC 1852
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7417 AGCATCTTGAACAAAGCATTTAAGCTTAACAAATTAACCAAAAGAAATGCC 7360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 6  
 US-09-438-142-1/c  
 Sequence 1, Application US/09438142  
 Patent No. 6258571

GENERAL INFORMATION:  
 APPLICANT: Ilya Chumakov  
 APPLICANT: Hiroaki Tanaka  
 TITLE OF INVENTION: High Throughput DNA sequencing vector  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe, Martens, Olson & Bear  
 STREET: 501 West Broadway  
 CITY: San Diego  
 STATE: California  
 COUNTRY: USA  
 ZIP: 92101-3505

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy Disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: win95  
 SOFTWARE: word

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/438.142  
 FILING DATE:

ATTORNEY/AGENT INFORMATION:  
 NAME: Israelson, Ned A.  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: GENSET. 015C1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10317 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double

TOPOLOGY: circular  
MOLECULE TYPE: synthetic DNA  
ORIGINAL SOURCE:  
ORGANISM: Cloning vector pGendEL  
FEATURE:  
NAME/KEY: pGendEL  
LOCATION: 1..10317  
FEATURE:  
NAME/KEY: Homology with X06404 compl (411..1668)  
LOCATION: 9..1266  
IDENTIFICATION METHOD: blastn against X06404  
FEATURE:  
NAME/KEY: Kanamycin resistance gene CDS  
LOCATION: 142..957  
IDENTIFICATION METHOD: By homology to X06404  
FEATURE:  
NAME/KEY: Tn1000's right end  
LOCATION: complement 1332..1371  
IDENTIFICATION METHOD: blastn against X06200)  
FEATURE:  
NAME/KEY: Homology with U46017 (1-472)  
LOCATION: 1423..1894  
IDENTIFICATION METHOD: blastn against U46017  
FEATURE:  
NAME/KEY: single stranded DNA replication origin  
LOCATION: 1423..1894  
IDENTIFICATION METHOD: By homology to U46017  
OTHER INFORMATION: mutation T -> C 1658  
FEATURE:  
NAME/KEY: Homology with U51113 (2382..6997)  
LOCATION: 1896..6544  
IDENTIFICATION METHOD: blastn against U51113  
FEATURE:  
NAME/KEY: OriS  
LOCATION: 1972..2188  
IDENTIFICATION METHOD: By homology to U51113  
FEATURE:  
NAME/KEY: repELR  
LOCATION: 2897..2918  
OTHER INFORMATION: Described in seqID 16  
FEATURE:  
NAME/KEY: RepE  
LOCATION: 2903..3034  
IDENTIFICATION METHOD: By homology to U51113  
FEATURE:  
NAME/KEY: T3  
LOCATION: 3043..3059  
OTHER INFORMATION: Described in seqID 17  
FEATURE:  
NAME/KEY: LRT3RA  
LOCATION: complement 3045..3069  
OTHER INFORMATION: Described in seqID 15  
FEATURE:  
NAME/KEY: IncC  
LOCATION: 3070..3320  
IDENTIFICATION METHOD: By homology to U51113  
OTHER INFORMATION: insertion 33 bases 3038..3071  
FEATURE:  
NAME/KEY: Para  
LOCATION: 3635..4821  
IDENTIFICATION METHOD: By homology to U51113  
OTHER INFORMATION: mutation G -> A 3878  
FEATURE:  
NAME/KEY: ParaB  
LOCATION: 4821..5792  
IDENTIFICATION METHOD: By homology to U51113  
FEATURE:  
NAME/KEY: ParC  
LOCATION: 5865..6382  
IDENTIFICATION METHOD: By homology to U51113  
FEATURE:  
NAME/KEY: Homology with J01688 (complement 175..819)  
LOCATION: 6574..7218

IDENTIFICATION METHOD: blastn against J01688  
OTHER INFORMATION: mutation A -> G 7096  
FEATURE:  
NAME/KEY: CDS streptomycin sensitivity gene  
LOCATION: complement 6716..7090  
IDENTIFICATION METHOD: By homology to J01688  
OTHER INFORMATION: mutation A -> G 6728  
OTHER INFORMATION: mutation G -> C 6821  
OTHER INFORMATION: mutation C -> T 6866  
OTHER INFORMATION: mutation T -> C 7013  
OTHER INFORMATION: mutation T -> A 7058  
FEATURE:  
NAME/KEY: fpsLR  
LOCATION: 7155..7174  
OTHER INFORMATION: Described in seqID 12  
FEATURE:  
NAME/KEY: SP6  
LOCATION: 7230..7248  
OTHER INFORMATION: Described in seqID 13  
FEATURE:  
NAME/KEY: Tn1000's left end  
LOCATION: 7252..7291  
IDENTIFICATION METHOD: blast (X06200)  
FEATURE:  
NAME/KEY: Homology with X02730 (complement 37..1959)  
LOCATION: 7305..9227  
IDENTIFICATION METHOD: blastn against X02730  
FEATURE:  
NAME/KEY: CDS levansucrase gene  
LOCATION: complement 7379..8800  
IDENTIFICATION METHOD: By homology to X02730  
OTHER INFORMATION: mutation T -> C 7466  
OTHER INFORMATION: mutation A -> G 7739  
OTHER INFORMATION: mutation T -> C (Asn -> Asp) 8347  
OTHER INFORMATION: mutation T -> C 8600  
OTHER INFORMATION: mutation G -> A (Ala -> Val) 8772  
FEATURE:  
NAME/KEY: SLR3  
LOCATION: 8711..8731  
OTHER INFORMATION: Described in seqID 14  
FEATURE:  
NAME/KEY: Homology with J01636 (complement 1158..1465)  
LOCATION: 9298..9623  
IDENTIFICATION METHOD: blastn against J01636  
FEATURE:  
NAME/KEY: CDS alpha peptide beta-galactosidase  
LOCATION: complement 9276..9497  
IDENTIFICATION METHOD: By homology to J01636  
FEATURE:  
NAME/KEY: primer HE1  
LOCATION: complement 9465..9479  
FEATURE:  
NAME/KEY: primer HE2  
LOCATION: 9461..9475  
FEATURE:  
NAME/KEY: primer lacIRS2Avr  
LOCATION: complement 9603..9630  
FEATURE:  
NAME/KEY: primer lacE2Mlu  
LOCATION: 9289..9314  
FEATURE:  
NAME/KEY: Homology with M77789 (1889..2576)  
LOCATION: 9629..10315  
IDENTIFICATION METHOD: blastn against M77789  
FEATURE:  
NAME/KEY: high copy-number double-stranded DNA replication origin  
LOCATION: complement 9629..10315  
IDENTIFICATION METHOD: By homology to M77789  
OTHER INFORMATION: mutation C -> T 9803  
OTHER INFORMATION: site Scal 10029 - 10034  
OTHER INFORMATION: site PmlI 10038 - 10043  
OTHER INFORMATION: CLONING SITES 10031 - 10041  
FEATURE:

```

NAME/KEY: oriLrd
LOCATION: 9856..9881
OTHER INFORMATION: Described in seqid 8
FEATURE:
NAME/KEY: OSI
LOCATION: 10009..10026
OTHER INFORMATION: Described in seqid 10
FEATURE:
NAME/KEY: ORI
LOCATION: complement 10046..10062
OTHER INFORMATION: Described in seqid 11
FEATURE:
NAME/KEY: oriLrd
LOCATION: complement 10182..10202
OTHER INFORMATION: Described in seqid 9
US-09-438-142-1

```

```

Query Match      25.2% Score 607.2; DB 4; Length 10317;
Best Local Similarity 66.3%; Pred. No. 1.5e-158;
Matches 993; Conservative 0; Mismatches 448; Indels 57; Gaps 6;

```

```

OY 358 ACGATGACCTTCAAAAGATGCGGCAAAAGAGCGCTGCGTAACTTCAGGACTGCTATA 417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8803 ACGATGACCTTCAAAAGATGCGGCAAAAGAGCGCTGCGTAACTTCAGGACTGCTATA 477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 418 TTATAGAGGAGCGGAGCGGCGCATATTTTGGCGCAAAATGACAGCGGAGCTACAG 477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8743 CTGGAGGAGG---CGCAACTCAACGCTTTGGCAAGAAAGCAACCAAGGCATATAG 8687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 478 GAGACTATGTTTGGCCATATTCAGCGGCTGACATGCTAAATTCAGAGACAA 537
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8686 GAAACATAGCGGATTTCCCATATTTACGCCATGATGCTGCAATCCCTGACAGCA 8627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 538 AACCTCCCATTAATAGTGCCTCATTCATGATGATGATGATGATGATGATGATG 597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8626 AAAAAATGAAAAATATCAAGTTCCTAGTTCATGCTCACAATTAATAATTCCTCTCT 8567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 598 GCAAAAGGATATGATAGTACAGCACTTAATAGATTAGATTGATGATGATGATG 657
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8566 GCAAAAGG-----CCTGGAGCTTTGGGAGACGTGGGCA 8534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 658 CTGCAAAAGCGTATGATGATGATGATGATGATGATGATGATGATGATGATG 717
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8533 TTACAAAAGCGTATGATGATGATGATGATGATGATGATGATGATGATGATG 8474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 718 GCAGGTGACCAAAAGAGTATGATGATGATGATGATGATGATGATGATGATG 777
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8473 GCCGAGATCCCTAAATATGCGGATGATGATGATGATGATGATGATGATGATG 8414
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 778 GATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 837
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8413 GAAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8354
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 838 GTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 897
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8353 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 898 ACCAAGATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 957
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8293 ACATGATGAGGAAAAATCCGTTATCTACATGATGATGATGATGATGATGATG 8241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 958 GGAACGGTGTGATTAACCAATCTTCACTGCTCAAGTAACTTCCAGCGGAT 1017
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8240 -----CGGCAAAACAACTGACACTGACAAAGTAAAGTATGACATCAGAC 8192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1018 GCAGTACACTTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1077
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8191 AGCTCT---TTGAACATACAGGCTAGAGATTAATCAATCATCTTTGA---CGGTAC 8138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1078 GTTACAGTTTCAAAATATTCAGCAATTTATGATGATGATGATGATGATGATGATG 1137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8137 GAAAAAGCTATCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATG 8078
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

OY 1138 AACCATCTTAAAGACCCCTACTATGTTGAGATAGAGCCATTAATATCTTGTCTT 1197
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8077 AACCATGCTGAGAGATCTCTACTAGTATAGATTAAGCCACCAATATCTTGTATTT 8018
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1198 GAAGCAATCTGAGACACAGATGTTATCAAGCGCATCAGTTTCAATTAATAGCT 1257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8017 GAAGCAAACTGAGACGATGATGATGATGATGATGATGATGATGATGATGATGATG 7958
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1258 TACTATGCGGAGGAGCTCTCTTCCAGAAATGAAAAAATTAAGTCTTCAAGTCT 1317
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7957 TACTATGCAAAACACATCATCTTCTTCCGTCAGAAAGTCAAAACTTCTGCAAGCAT 7898
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1318 AAAAAACAATTTGCTTTTAAAGCAATGTCATTAAGCATTTGTAATTTGGCGATGAC 1377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7897 AAAAAAGCAGCGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 7838
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1378 TATACAGTGAAGTGTATTAACCATTAAGTCTGATCAACACAGTATGATGATGATG 1437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7837 TACACACTGAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 7778
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1438 GAAGCGCCCATTAATTAATGATTAATTAATGATGATGATGATGATGATGATGATG 1497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7777 GAAGCGCGGAGCGTCTTAAATGACGCAAAATGATGATGATGATGATGATGATGATG 7718
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1498 TCAAAATGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7717 TCAAAATGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 7658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1558 GACTCTTAATTTGCGCCACACACCCGATTAATTAATGATGATGATGATGATGATG 1617
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7657 AATCTTTAACTGCGCCATCAACACCGCTGAAACAACTGCGCTTGTGTTAAAAATGAT 7598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1618 CTGACCCCTGCTGATCTCACACACCTTCTGATGATGATGATGATGATGATGATG 1677
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7597 CTGATCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1678 AATAATGTGATCTACAAATTAATGAGATGATGATGATGATGATGATGATGATG 1737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7537 AACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1738 CACCTGCGGAGCAAGCTTGGGGTTAATTTAAAGGTCGTGACATCTGAGGAGAAAT 1797
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7477 ACGTTTGGCGGAGCTTCTCTGACATCAACGCAAGCAAGAAACATGTTGTCAAGAC 7418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1798 AGTTCC---GGACAAAGCAATTCCTCAGATGATGATGATGATGATGATGATG 1852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7417 AGCATCTTGAACAAGCATTAACGTAAACAATTAACCAAGCAAGAAAGAAATGCC 7360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 7
US-09-068-043-1
; Sequence 1, Application us/09068043
; Patent No. 6048694
;
GENERAL INFORMATION:
APPLICANT: MICHAEL GENE BRAMUCCI
APPLICANT: VASANTHA NGARAJAN
TITLE OF INVENTION: POSITIVE SELECTION
TITLE OF INVENTION: VECTOR FOR BACILLUS SP.
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND
ADDRESSEE: COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT OFFICE 97
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/09/068,043
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/006,201
FILING DATE: NOVEMBER 3, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FELTHAM, S. NEIL
REGISTRATION NUMBER: 36,506
REFERENCE/DOCKET NUMBER: CR-9807
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-6460
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3305 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
US-09-068-043-1

```

```

Query Match      23.0%; Score 553.8; DB 3; Length 3305;
Best Local Similarity 64.6%; Pred. No. 6,1e-144;
Matches 932; Conservative 0; Mismatches 457; Indels 54; Gaps 5;

```

```

QY 357 GACGATGACCTTCAAAAGATTGGCGAAAAACACACTGCCCTACCTCAGACACTGCTAT 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1801 GAACATGACATCAAAATTTGTAAAGACACAGTACGACTAGCTTACCTGCACT 1860

QY 417 ATTACTAGAGCGAGCGACCGCATATTTTGGCGCAAAATGACAGCGGGGACTACAA 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1861 GCTACACAGA---GGAGCACTCAGCCTTCTGGAAAGAAATTAACCAAAAGCATACAA 1917

QY 477 GGAAGACTATGTTTGGCCATATTACACGCGCTGACATGCTAAAAATTCAGACAACA 536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1918 AGAAACGTACGCGCTCTCATATTACACGCCCATGATATGTCGAGATCCCTAAACAGA 1977

QY 537 AAACGCTCTCAATTTAAATGCTCAATTCATGATGATGATGATGATGATGATGATG 596
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1978 GCATAACGAAATATCCAAAGTGCCTCAATTCATGATGATGATGATGATGATGATG 2037

QY 597 GGCATAAGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 656
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2038 TGCAAAAG-----ACTGATGCTGCTCGACAGCTGCGC 2070

QY 657 ACTGCAAAACGCTGATGCTGCTGCGCAAAATTCATGATGATGATGATGATGATG 716
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2071 GCTGCAAAACGCTGACGAGACAGTACGAAATACAAAGGCTATGCTGCTGCTGCT 2130

QY 717 AGCAGGTACCCCAAAACAGTATGATGATGATGATGATGATGATGATGATGATG 776
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2131 TGCGGGAAGCCCGAAGAGCGATGACATCATCTATGATGATGATGATGATGATG 2190

QY 777 TGATACATCATGATGACAGCTGGAAGATGATGATGATGATGATGATGATGATG 836
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2191 CGACAATCATGACAGCTGGAAGAGCGCGCTGCTTAAAGACAGGATGATGATG 2250

QY 837 TGTTCATATGATGCTGCTGCTTAAATATCAACACAGAGTGTGAGGTCTGCTACTTT 896
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2251 CGAGCCCAACGATCGATCTGGAAGATGACAGCAAGATGCTCGGTTGCAACCTT 2310

QY 897 AACCAAGATGGCCAGTCCGTTTATTCATACAGTATGATGATGATGATGATGATG 956
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2311 TACATCTACGGAAGAAATCCGTTTATCTACACTACTATCTCGGTAACATTA----- 2364

QY 957 TGAACCGGCTGTGTAACCAATCTTCACTGCTCAAGTAATCTTATCCAGCCGGA 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2365 -----CGCAAAACAAAGCTGCAACAGCAGGATGATGATGATGATGATGATG 2412

QY 1017 TGCAGTCACTTAAAGTCAGTATGATGATGATGATGATGATGATGATGATGATG 1076
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 2413 T--GACACACTCAAAATCAACGAGTGAAGATCACAAAACGATTTTGA---CGAGA 2466

QY 1077 CGGTACAGTTTATCAAAATATTCACCAATTTATGATGAAAGCAGTGGATTTACGTA 1136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2467 CGGAAACATATACAAAGCTTACAGTGTATATGATGAAAGCAATTTATATTCGCGGA 2526

QY 1137 TAACCATCTTTAAGACCCCTCAGTATGTAAGATTAAGGCCATTAATATCTGTGTT 1196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2527 CAACCATACGCTGAGAGACCCCTCAGTATGTAAGATTAAGGCCATTAATATCTGTAT 2586

QY 1197 TGAACGGAATCTGGAACACAGATGTTATCAAGGCGATGAGTCTTCAATATTAAGC 1256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2587 CGAAGCCAAACAGGAAACAGAAACGGATACCAAGGCGAAGATCTTTATTAACAAAGC 2646

QY 1257 TTACTATGGCGGAAGTGTGCTTCTTCCAGATGAAAAAATTAACCTCTTCAAGTCC 1316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2647 GTACTACGGCGGCGACGACACTTCTCGTAAAGAAACGAGAACTTACACAGAGCC 2706

QY 1317 TAAAAAACAAATTCCTTTTATGCAATGATGATGATGATGATGATGATGATGATG 1376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2707 TAAAAAACCGCATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2766

QY 1377 CTATACAGTGAAGAGTATGAAACCATTTAGTGCATCAACACAGATGATGATGATG 1436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2767 TTACACATGAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2826

QY 1437 GGAAGCGGCAATATTTAAATGATATTAATGATGATGATGATGATGATGATGATG 1496
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2827 CGAGCGCGGGAATGTTTCAAAATGATGATGATGATGATGATGATGATGATGATG 2886

QY 1497 ATCCAAATGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1556
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2887 TTCAAAATGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2946

QY 1557 GCATCTCTTAAATGGCCACACAAACCGATTAATGAAGTGAAGTGAATGAAGTGA 1616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2947 AAACCTTTAAACCGCCCTTACAAAGCGCTGAACAAAGGAGGCTGCTCAATGAG 3006

QY 1617 TCTTGACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3007 TCTTGATTCAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3066

QY 1677 TAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1736
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3067 CAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3126

QY 1737 TCACCTGGGGAACAAGCTTGGGTTAATTAAGGCTGACACATCTGAGAGGAAAA 1796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3127 AACATTTGGCCCAAGCTTCTTAATCAACATCAAAAGCAATTAACATCGTTGTCAAAA 3186

QY 1797 TNG 1799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3187 CAG 3189

```

```

RESULT 8
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/232,463  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZgpt-F15  
US-08-232-463-14

Query Match 1.9%; Score 44.8; DB 1; Length 7218;  
Best Local Similarity 8.8%; Pred. No. 0.02;  
Matches 40; Conservative 211; Mismatches 203; Indels 0; Gaps 0;  
QY 289 CTATATCTTTTATATAGAACTATCAAGAGCTTCTTATCAATTCATACATATCCA 348  
DB 1475 CTATCTAGCAAGTGTAAAGATAGAGATTTGGTACRRRRRRRRRRRRRRRR 1416  
QY 349 AGGAGGAGAGATCACTCAAAAGATTGGGAAAAGCGCTGCGTAACCTCAG 408  
DB 1415 RRR 1356  
QY 409 ACTGCTATATAGTAGAGGAGCGACGCATATTTTGGCCAGCAATGAAACGCGG 468  
DB 1355 RRR 1296  
QY 469 GACTCAAGGAAGACTATGTTTCCCATATTAACGCGCTGACATCTTAAATTCGA 528  
DB 1295 RRR 1236  
QY 529 GGACAAACAAACAGCTCAATTTAAAGTGCCTCAATTCATGATCAGCAATCAAAAC 588  
DB 1235 RRR 1176  
QY 589 ATTGATTCGGCAAAAGGTATGATAGTCAGCAACTAATAGATTAGATGATGATG 648  
DB 1175 RRR 1116  
QY 649 AGCTGGCACTGCAGAAAGCTGATGCTACTGCGCAAAATTATCATGATCATCTGC 708  
DB 1115 RRR 1056  
QY 709 TCCGCTTAGCAGGTCAGCAAAACAGATG 742  
DB 1055 CCTGAGCTGCAAGCAAGCTCGGAATTATCTG 1022

RESULT 9  
US-08-671-947-1  
Sequence 1, Application US/08671947  
Patent No. 5827515  
GENERAL INFORMATION:

APPLICANT: Shahabi Reynoso, Mitra  
APPLICANT: Yamamoto, Takashi  
APPLICANT: Cooper, Nicole H.  
APPLICANT: Kalman, Sue S.  
TITLE OF INVENTION: BACILLUS THURINGIENSIS SPOROLATION GENE  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SANDOZ AGRO, INC.  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/671,947  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Marcus-Wyner, Lynn  
REGISTRATION NUMBER: 34,869  
REFERENCE/DOCKET NUMBER: 133-0724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/354-3588  
TELEFAX: 415/857-1125  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1662 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 474..1427  
OTHER INFORMATION: /codon\_start= 474  
US-08-671-947-1

Query Match 1.7%; Score 39.8; DB 1; Length 1662;  
Best Local Similarity 67.5%; Pred. No. 0.24;  
Matches 56; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 1452 ATTTAAATGATATATAATGATCTATTCACGAGATTCAGAGATCCAAATGACGAG 1511  
DB 652 ATATATAAAGATATATAAAGAAATATATCTGATTTATGTAATAAAAAGACAAAG 711  
QY 1512 TGATGAAATTAAAGCAACAAGATG 1534  
DB 712 AGAAGGATTTGAGTCAGAGAAG 734

RESULT 10  
US-09-276-531-82/C  
Sequence 82, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lynn E. Murry, Ph.D.  
REGISTRATION NUMBER: 42,918  
REFERENCE/DOCKET NUMBER: PA-0008 US  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 807 base pairs  
TYPE: nucleic acid  
STRANDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRAITUT12  
CLONE: 1617155  
US-09-276-531-82

Query Match 1.6%; Score 38; DB 4; Length 807;  
Best Local Similarity 57.9%; Pred. No. 0.55;  
Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1853 ACTGGCAAGATCTTACCTTTTCCCAACCTTTTAAATAAGGGTTTGTACAC 1912  
Db 250 ATTCCCAATTTCACTGGGGCCCGCTTTTAAANAAAGGCTCGTGAAC 191

Qy 1913 CTTAGAAGAAAGATCTTGGCCCGGCAATTACCATGCC 1959  
Db 190 CTTGGGGGAAANCCCTTGGCGGGTTAANCCCAACC 144

RESULT 11  
US-09-134-001C-650  
Sequence 650, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 650  
LENGTH: 1434  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-650

Query Match 1.6%; Score 37.6; DB 4; Length 1434;

Best Local Similarity 51.8%; Pred. No. 0.93;  
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 84 ACAAACAGCTTACGATGACCAATTAAGAAATTTCAAAAGTCTGGGATAGGCC 143  
Db 732 AGCCATTATCATATTTTACACGCGTCACTTAATAAAGATCATGTTAATAT 791

Qy 144 AAGTCACTTTGAATAAATTAAGCTCTGTCGGGGCTTTTGTATTATTTAT 203  
Db 792 GCGTGTGTTAAATAAATTAAGTGTTCACACTTTTCAGCTGTTTCAATGCTGTGTAATAT 851

Qy 204 TTCAACTGCAAGTGTCCATCCCTATATCATTTTAAGACGAA 247  
Db 852 TACGATGTAAGTCCGCTTACATCATATTTACATTCAAA 895

RESULT 12  
US-08-480-604A-9  
Sequence 9, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 42A  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7101 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..7098  
US-08-480-604A-9

Query Match 1.5%; Score 36.6; DB 1; Length 7101;  
Best Local Similarity 60.6%; Pred. No. 3.7;  
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 622 AACTTAATAGATTGATGTAGTGGCCACTGCAAAACGCTGATGCTACGCG 681  
DB 3478 AATTCATAAGTTTGTAGTAAATGTGAATCTGAGAAATGGAAGGTGTTCAAGTCATACT 3537

OY 682 GCAATTTATCATGATATCATCATCGTCCTCCGCTTAGCA 720  
DB 3538 GTAACGTGATGATATGATGATCACTTCTTTTCAGCACCATCA 3576

RESULT 13  
US-08-405-496A-9  
Sequence 9, Application US/08405496A  
Patent No. 5919665  
GENERAL INFORMATION:  
APPLICANT: WILLIAMS, JAMES A.  
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
TITLE OF INVENTION: NEUROTOXIN  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,496A  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7101 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..7098  
US-08-405-496A-9

Query Match 1.5%; Score 36.6; DB 2; Length 7101;  
Best Local Similarity 60.6%; Pred. No. 3.7;  
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 622 AACTTAATAGATTGATGTAGTGGCCACTGCAAAACGCTGATGCTACGCG 681  
DB 3478 AATTCATAAGTTTGTAGTAAATGTGAATCTGAGAAATGGAAGGTGTTCAAGTCATACT 3537

OY 682 GCAATTTATCATGATATCATCATCGTCCTCCGCTTAGCA 720  
DB 3538 GTAACGTGATGATATGATGATCACTTCTTTTCAGCACCATCA 3576

RESULT 14  
US-08-915-136-9  
Sequence 9, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FTRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 9:







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 23:20:16 ; Search time 531.636 Seconds  
(without alignments)  
10200.228 Million cell updates/sec

Title: US-09-986-682B-5  
2408  
Perfect score: 1 CGGGGAAATACTAGATTCC.....TACCAATMAAATGTCATC 2408  
Sequence:

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2408	100.0	2408	19	AAV17621
2	627	26.0	1449	24	ABK75353
3	607.2	25.2	10317	21	AAZ53988
4	577.8	24.0	1668	21	AAZ57329
5	577.8	24.0	1722	21	AAZ57330
6	564.2	23.4	2568	23	AAS68687
7	563.4	23.4	1584	23	AAS68494
8	563.4	23.4	2044	23	AAS85834
9	563.4	23.4	2044	23	AAS87384

10	563.4	23.4	2325	23	AAS69777	DNA encoding novel
11	563.4	23.4	2325	23	AAS75941	DNA encoding novel
12	563.4	23.4	2351	12	AAO10498	B. amylioliquefacien
13	563.4	23.4	2370	23	AAS69133	DNA encoding novel
14	563.4	23.4	2370	23	AAS92594	DNA encoding novel
15	563.4	23.4	2374	23	AAS87392	DNA encoding novel
16	563.4	23.4	2472	23	AAS89956	DNA encoding novel
17	563.4	23.4	2479	23	AAS86776	DNA encoding novel
18	563.4	23.4	2682	23	AAS71354	DNA encoding novel
19	563.4	23.4	2682	23	AAS87361	DNA encoding novel
20	563.4	23.4	3084	23	AAS85832	DNA encoding novel
21	563.4	23.4	3084	23	AAS87383	DNA encoding novel
22	563.4	23.4	3505	23	AAS85838	DNA encoding novel
23	563.4	23.4	4132	23	AAS86045	DNA encoding novel
24	563.4	23.4	6721	24	AAS18599	Purineergic recepto
25	563.4	23.4	6721	24	AAS18600	Purineergic recepto
26	561.8	23.3	1690	23	AAS85792	DNA encoding novel
27	561.8	23.3	1690	23	AAS86019	DNA encoding novel
28	561.8	23.3	1690	23	AAS87350	DNA encoding novel
29	561.8	23.3	3288	23	AAS77395	DNA encoding novel
30	561.8	23.3	3288	23	AAS85789	DNA encoding novel
31	561.8	23.3	3288	23	AAS86010	DNA encoding novel
32	561.8	23.3	3288	23	AAS87346	DNA encoding novel
33	561.8	23.3	3288	23	AAS88746	DNA encoding novel
34	555.6	23.1	3238	23	AAS64228	DNA encoding novel
35	553.8	23.0	3305	18	AAT68806	Mutant levansucras
36	547.4	22.7	1935	23	AAS70417	DNA encoding novel
37	547.4	22.7	2562	23	AAS88281	DNA encoding novel
38	547.4	22.7	2742	23	AAS66648	DNA encoding novel
39	547.4	22.7	2790	23	AAS87386	DNA encoding novel
40	547.4	22.7	3267	23	AAS85793	DNA encoding novel
41	547.4	22.7	5912	23	AAS86048	DNA encoding novel
42	547.4	22.7	5938	23	AAS87395	DNA encoding novel
43	545	22.6	3086	23	AAS86052	DNA encoding novel
44	532.8	22.1	1629	23	AAS85800	DNA encoding novel
45	532.8	22.1	1629	23	AAS86022	DNA encoding novel

ALIGNMENTS

RESULT 1	
AAV17621	AAV17621 standard; DNA; 2408 BP.
ID	AAV17621
AC	AAV17621:
DT	04-JUN-1998 (first entry)
DE	Bacillus sp. V230 beta-fructofuranosidase genomic DNA.
XX	
KW	Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
KW	lactosucrose; fructosyltrehalose; antidiabetic; growth promoter;
KW	Bifidobacteria; mineral absorption promoting activity; taste; texture;
KW	cosmetic; pharmaceutical; sweetener; ds.
OS	Bacillus sp.
FH	
FT	Key
FT	5'UTR
FT	Location/Qualifiers
FT	1..360
FT	/*tag= a
FT	sig_peptide
FT	361..456
FT	/*tag= b
FT	mat_peptide
FT	457..1821
FT	/*tag= c
FT	/EC_number= 3.2.1.26
FT	/product= beta-fructofuranosidase
FT	/note= "Claim 7"
FT	1822..2408
FT	/*tag= d
FT	3'UTR
FT	
XX	
XX	EP812915-A2.
XX	

PD	17-DEC-1997.
PF	10-JUN-1997; 97EP-0304032.
PR	10-JUN-1996; 96JP-0170630.
XX	(HAYB ) HAYASHIBARA SEIBUTSU.KAGAKU.
PA	Chen H, Kubota M, Tsusaki K;
PI	WPI: 1998-034976/04.
DR	P-PSDB: AAM48306.
XX	Bacillus beta-fructofuranosidase enzyme - useful for producing
PT	fructo-furanosylated saccharide(s) or alcohol(s)
XX	
PS	Claim 8; Pages 23-25; 32pp; English.
XX	
CC	This is a genomic DNA sequence which encodes
CC	beta-fructofuranosidase, an enzyme that has an optimum temperature
CC	of 50 deg. C at pH 6. The enzyme was produced by recombinant DNA
CC	technology. It catalyses fructofuranosyl transfer reactions from a
CC	fructofuranosyl donor to a fructofuranosyl acceptor where the donor is
CC	sucrose, raffinose or elose and the acceptor is selected from alcohols,
CC	sugar alcohols and saccharides having no beta-fructofuranosidic linkages,
CC	especially where the reaction product is xylosyl fructose, elose,
CC	isomaltosyl fructose, lactosucrose or fructosylltrehalose.
CC	Such reaction products have a satisfactory taste and sweetness, a
CC	moderate viscosity and humectancy, an effective anticariogenic activity,
CC	growth promoting activity for Bifidobacteria, mineral-absorption
CC	promoting activity. The products can be used to improve the tastes and
CC	textures of food products, cosmetics and pharmaceuticals and are useful
CC	as sweeteners in the food and pharmaceutical industries.
XX	
SO	Sequence 2408 BP: 777 A; 492 G; 496 G; 643 T; 0 other:
Query Match	100.0%; Score 2408; DB 19; Length 2408;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2408; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	1 CGGGGAAATACTAGATTCCAATTTGGCCAGACTTCCCAGTTGGTGTAAGAAGAGTTGC 60
Db	1 CGGGGAAATACTAGATTCCAATTTGGCCAGACTTCCCAGTTGGTGTAAGAAGAGTTGC 60
OY	61 GACGCCAATGCAGCTGTGGCGTAGAAAAACAGCTTACTCATGACAATTTAGGAAGAT 120
Db	61 GACGCCAATGCAGCTGTGGCGTAGAAAAACAGCTTACTCATGACAATTTAGGAAGAT 120
OY	121 TTCAAAAGTCGGGATGAAGGCCAACAGTCCACTTTGAANAATTTTCAGCCTCGTGTG 180
Db	121 TTCAAAAGTCGGGATGAAGGCCAACAGTCCACTTTGAANAATTTTCAGCCTCGTGTG 180
OY	181 GGGGCTTTTTGTTTTTATTTATTTCAACTGCAAGTGGTCATCCCTATATCAATTAA 240
Db	181 GGGGCTTTTTGTTTTTATTTATTTCAACTGCAAGTGGTCATCCCTATATCAATTAA 240
OY	241 GACGAAATTTCAATCAATCCAGCATCCCCCAATAACGCGTCCTCTATACCTTTAA 300
Db	241 GACGAAATTTCAATCAATCCAGCATCCCCCAATAACGCGTCCTCTATACCTTTAA 300
OY	301 TTAATTAAGAACTATCAAGAGCTTCTTATCAAAATTCATACATTCACAAGAGGAGAGC 360
Db	301 TTAATTAAGAACTATCAAGAGCTTCTTATCAAAATTCATACATTCACAAGAGGAGAGC 360
OY	361 ATGAACCTTCAAAAAGTTGGCGAAAAAGCAGCTGCCGTAACCTTACAGACTCTATATTA 420
Db	361 ATGAACCTTCAAAAAGTTGGCGAAAAAGCAGCTGCCGTAACCTTACAGACTCTATATTA 420
OY	421 GTAGAGAGGAGAGCAGCCGATATTTTTGGCGAGCAAATTAACGCGGGAGATCAAGAGAA 480
Db	421 GTAGAGAGGAGAGCAGCCGATATTTTTGGCGAGCAAATTAACGCGGGAGATCAAGAGAA 480
OY	481 GACTATGCTTTTGCCCATATTAACACGCGCTGACATGCTAAAAATTCACAGACAAAMAAC 540

Db	481	GA	CTATGGTTTGGCCATTATACAGCGCGTACATGCTAATAAATTCAGAGCAACAAAC	540
Qy	541	AG	CTCCATTTAAAGTGCCTCAATTCATGCATCAGCAATCAAAAACATTGATTCGGCA	600
Db	541	AG	CTCCATTTAAAGTGCCTCAATTCATGCATCAGCAATCAAAAACATTGATTCGGCA	600
Qy	601	AA	AGGTATGATTAAGTCAAGGCACTTATACATTTAAGATGATATGGGATAGCTGGCCACTG	660
Db	601	AA	AGGTATGATTAAGTCAAGGCACTTATACATTTAAGATGATATGGGATAGCTGGCCACTG	660
Qy	661	CA	AAACGCGTATGATGACGCGGCAAAATATCATGGATATCATCATCGTCCCGTTTAGCA	720
Db	661	CA	AAACGCGTATGATGACGCGGCAAAATATCATGGATATCATCATCGTCCCGTTTAGCA	720
Qy	721	GG	TGACCCAAAAAACAGTATGATCTCCACTCATTTATTTCTATCAAAAAGTCGGTGAT	780
Db	721	GG	TGACCCAAAAAACAGTATGATCTCCACTCATTTATTTCTATCAAAAAGTCGGTGAT	780
Qy	781	AC	ATCGATTTGACAGCTGGAAAAATGCGSAGATATTTGAAATGAAATGATGAAATTTGTT	840
Db	781	AC	ATCGATTTGACAGCTGGAAAAATGCGSAGATATTTGAAATGAAATGATGAAATTTGTT	840
Qy	841	CC	AAATGATCCGTATCTTAAATATCAACACAGAGAGTGCTAGGTTCTGCTACTTTAAC	900
Db	841	CC	AAATGATCCGTATCTTAAATATCAACACAGAGAGTGCTAGGTTCTGCTACTTTAAC	900
Qy	901	AA	AGATGGCCAAAGTCGGTTATTCATATCAATTAACAGGATATCCGTAAGATGATGGA	960
Db	901	AA	AGATGGCCAAAGTCGGTTATTCATATCAATTAACAGGATATCCGTAAGATGATGGA	960
Qy	961	ACC	GTGCTGGTAAACCAATCATTTTCAACCTGCTCAAGTAACTTATCCACCGCGATGCA	1020
Db	961	ACC	GTGCTGGTAAACCAATCATTTTCAACCTGCTCAAGTAACTTATCCACCGCGATGCA	1020
Qy	1021	GCT	ACACTTAAAGTCAGTATCTATCATATAATCTGCTTGTGATGGCGGAGACGCT	1080
Db	1021	GCT	ACACTTAAAGTCAGTATCTATCATATAATCTGCTTGTGATGGCGGAGACGCT	1080
Qy	1081	AC	AGTTTATCAAAATATTCACCAATTTATGATGAAAGGCAAGTGGATTTACAGTATAAC	1140
Db	1081	AC	AGTTTATCAAAATATTCACCAATTTATGATGAAAGGCAAGTGGATTTACAGTATAAC	1140
Qy	1141	CAT	ACTTTAAGAGACCCCTCATTATGTTCAGATTAAGGGCCATAATATCTTGTCTTTGAA	1200
Db	1141	CAT	ACTTTAAGAGACCCCTCATTATGTTCAGATTAAGGGCCATAATATCTTGTCTTTGAA	1200
Qy	1201	GCG	ATACTGGAACACAGATGTTATCAAGGCGATCGCTTTCAAATAAATAAAGCTTAC	1260
Db	1201	GCG	ATACTGGAACACAGATGTTATCAAGGCGATCGCTTTCAAATAAATAAAGCTTAC	1260
Qy	1261	TAT	GCGGAGAGTACGCTCTTCACGATCAAAAAATAACGCTTCAAAAGTCCTTAA	1320
Db	1261	TAT	GCGGAGAGTACGCTCTTCACGATCAAAAAATAACGCTTCAAAAGTCCTTAA	1320
Qy	1321	AA	CAAAATTTGCTTTTGAAGGATGTCATTTAGGATTTGAAATTTGGCGGATGACAT	1380
Db	1321	AA	CAAAATTTGCTTTTGAAGGATGTCATTTAGGATTTGAAATTTGGCGGATGACAT	1380
Qy	1381	AC	AGTGAAGAAGTGTATGAAACCATTTAGCATCAACACAGTATGAGTGAAGTGC	1440
Db	1381	AC	AGTGAAGAAGTGTATGAAACCATTTAGCATCAACACAGTATGAGTGAAGTGC	1440
Qy	1441	CG	CGGCATATATTTAAATGAATTAATTAATGATATCTATTCACAGGATTTCAAGAGATCC	1500
Db	1441	CG	CGGCATATATTTAAATGAATTAATTAATGATATCTATTCACAGGATTTCAAGAGATCC	1500
Qy	1501	AA	ATAGCAGATGATGAAATTAACGACAAAGATGTTTATGCTATGAGGCCCGGAGCGAC	1560
Db	1501	AA	ATAGCAGATGATGAAATTAACGACAAAGATGTTTATGCTATGAGGCCCGGAGCGAC	1560
Qy	1561	TC	CTTAAATGGCCACACACCGATTAATGAACATGACTTGTATTTGACATGATCTT	1620
Db	1561	TC	CTTAAATGGCCACACACCGATTAATGAACATGACTTGTATTTGACATGATCTT	1620



```

Db      232  AAGGGTATTAACAAAATGAGAGAGCTGATTCATTTAGAGATATGGAGACAGTCCGCTG 291
OY      661  CAAAACGCTGATGCTGCGGCAATATATCATGATATACATGCTCCGCTTTACGA 720
Db      232  CAAAATGCGCAGGAGGCTTGTCTACATACACAGCTCAAACTTGTGTTTCCGCTGGCG 351
OY      721  GGTGACCCCAAAAACAGTATGATCTCCACTTCTTATTTATCTATCAAAAAGTCGAT 780
Db      352  GCGCATTCGGAAGAGAGCTGATGACATCCATCTTTTGTCTATCAAAAAGAGCGCA 411
OY      781  ACATGATTTGACAGCTGGAATAATCTGAGAGATTTGAAGATATGATTAATTTGTT 840
Db      412  ACTTCTATCGACAGCTGGAATAAACCCGCGACAGATGTTTAAAGACAGCGCAAAATTTGT 471
OY      841  CCAATGATCCGCTATCTTAATATCAACACAGAGTGCTGAGTCTGCTCTTAAAC 900
Db      472  CCAGACGATCCGCTACCTCAAAACATTAACACAGGATGCTGAGTTGCTCCAGCTGCA 531
OY      901  AAAAGTGGCCAAAGCTGTTTATTTATCTATACAGATTTACTAGTAAATCTGAAAGTGTGA 960
Db      532  AAAAGCGGAAAAGTCCGACTGTTTACAAAGCTTTTCCGCGACGCAATA----- 581
OY      961  ACCGCTGCTGTATACCAATCTTTCAACTGCTCAAGTAACTTATCCAGCGGATGCA 1020
Db      582  -----CGGCAACACAGACGCTGACAAACAGCTCAGTCAATTTCTCAGCGGATTCG 633
OY      1021  GCTACACTTAAAGTGTAGATGATCTGATATCATTAATCTGTTTGAAGGCGGAGACGCT 1080
Db      634  GACACGCTCAAAATATGACGCTGTACAAAGATCATTAATCGGCTTTGACGCGCGCACGCG 693
OY      1081  ACAGTTATCAAAATATTTAGCAATTTATCGATGAGCGCAAGTGAATTCAGGTATAC 1140
Db      694  ACGGTATACCAAAAGCTTACAGCAATTCATTCAGCAAGAAACTACAGCTCCGCGCACAC 753
OY      1141  CATACTTTAAGAGACCTTCACATATTTGAAGTAAGGCGCAATATCTTCTTGA 1200
Db      754  CATACGATGAGAGACCCGCTATATGTGAAGACCGCGCAATTAATCTCTATTTGAA 813
OY      1201  GCGAATCTGCAACACAGATGTTATCATAGGCGATCATCTTCTTAATTAATTAAGCTTAC 1260
Db      814  GGCATACGGGAGCAAAAACCGGCTACCAAGAGAAAGACTCCCTATTCACAGAGACCTAC 873
OY      1261  TATGCGGCAAGTACAGCTCTCTCCAGATGAAAAAATAAATCTCTCAAGTCTTAA 1320
Db      874  TACGGGGGACACCAAAAGTCTTTTAAAGAAAGAACACAGCAAGCTCTCAAGAGTCCGAC 933
OY      1321  AAACAATTTGCTCTTTAGCCAGATGCTCATTTAGCATTTGGAATGGCCATACAT 1380
Db      934  AAAAAGAACGCTTCCGCTGCTACAGCGCTCTCGGAATTCATGAAATTAATTAACGATTA 993
OY      1381  ACAGTGAAGAAGTGTATGAACCATTTAGTCGATCAAAACAGAGTGAAGTAACTGCA 1440
Db      994  ACACGAAAAAAGTATGATAGAGCTTTGATCGCTCCATACGCTGACAGATTAATTCGA 1053
OY      1441  CGCGCCAAATATTTAAATGAATTAATTAATGATCTATTCACGAGTTCAGAGAGATCC 1500
Db      1054  CGGGCCAACTCTTCAAAATGAATGAATAATGATCTGTCATCGATTCAGATTCAGAGATCA 1113
OY      1501  AAAATGAGAGATGATGATTAACACAAAGATGTTTATGCTAGGCGCGCGAGCGCAC 1560
Db      1114  AAAATGACAATTTGAGCGATCGTTCAAAAGACATTTATGCTGCTGATGATTAACAGT 1173
OY      1561  TCTTTAATGCGCCACACAAACCGGATTAATGAATGAGAGTGTATGATGAATCAATCTT 1620
Db      1174  TCATTAAACCGGACCTTCAAGCTTTAAACAATCCGAGCTTTGTTTGCATATGAGACAG 1233
OY      1621  GACCTGCTGATCTCACACACATTTCTCATTTGCGATATCCGACCTGAGAGTAA 1680
Db      1234  GATTACATGACATACGCTTTACTTATTCACACTTTGCCGTAACCGCAAAAAAGCGCAC 1293
OY      1681  AATGTGATCTCACAGATTAATGAGATAGAGCTTCTATCCGAACATCATCTCTAC 1740
Db      1294  GAAGTGTATTAACAAGCTACATCAAAACAGAGGATTTGAAAGACATCATCCGACG 1353

```

```

OY      1741  CTGCGGACAAGCTTGGGTTAATTAAGGCTGACACATCTGAGAGAAAAATAG 1799
Db      1354  TTTGCACCAAGCTTTTCTCTGACATCAAAAGATCAAAACATCCGTTGTCAAAAACAG 1412

RESULT 3.
AA235988/C
ID      AA235988 standard: DNA: 10317 BP.
XX
XX      AA235988;
XX
DE      09-FEB-2000 (first entry)
XX
XX      Vector pGemb1 sequence SEQ ID NO:1.
XX
XX      DNA sequencing vector: nested deletion; transposition: contig;
XX      truncated lacZ; origin of replication: pGemb1; mapping: ss.
XX
XX      Synthetic.
XX
XX      WO9953044-A2.
XX
XX      21-OCT-1999.
XX
XX      09-APR-1999; 99WO-IB00690.
XX
XX      10-APR-1998; 98US-0058746.
XX
XX      (GEST ) GENSET.
XX
XX      Chumakov I, Tanaka H;
XX
XX      WPI: 2000-023040/02.
XX
XX      New DNA sequencing vectors, used for sequencing large regions of DNA
XX      PT and mapping the location of markers
XX
XX      Claim 43; Page 93-96; 102pp; English.
XX
XX      The present invention describes a vector comprising the following:
XX      CC (a) a high copy number origin of replication (Ori) having at least one
XX      CC cloning site, with at least one cloning site being positioned in the
XX      CC high copy number Ori such that the ability of the high copy number Ori
XX      CC to direct replication is not disrupted when no insert has been cloned
XX      CC into the cloning site and is disrupted when an insert is cloned into
XX      CC the cloning site; and (b) a low copy number Ori. Also described is a
XX      CC truncated lacZ gene to distinguish cells with the truncated lacZ
XX      CC gene at high copy number from cells carrying it at a low copy number,
XX      CC where the former are dark blue when grown on medium containing Xgal and
XX      CC IPNG and the latter are light blue when grown on medium containing Xgal
XX      CC and IPNG. Products and methods from the present invention can be used
XX      CC for the manipulation of DNA. They can be used for sequencing large
XX      CC regions of DNA and mapping the locations of markers within large regions
XX      CC of DNA. The present sequence represent a specifically claimed vector
XX      sequence from the present invention.
XX
XX      Sequence 10317 BP; 2567 A; 2296 C; 2469 G; 2985 T; 0 other;
XX
XX      Query Match 25.2%; Score 607.2; DB 21; Length 10317;
XX      Best Local Similarity 66.3%; Pred. No. 6.2e-161;
XX      Matches 993; Conservative 0; Mismatches 448; Indels 57; Gaps 6;

OY      358  ACGATGACCTTCAAAAGATTTGGCGAAAAACAGCTGCGGTAACCTTACAGACGCTATA 417
Db      8803  ACGATGACATCAAAAGATTTGCAAAACAAATACAGATTAACCTTTATCAACGCGACTG 8744
OY      418  TTATGAGAGCGGAGCGACCGCATATTTTTCGCGAGCAATGAACAGCGGAGTACAG 477
Db      8743  CTGCGCAGAGG---CGCACTCAAGCGTTTTCGGAAGAAAGCAACCAAAAGCCATATAG 8687
OY      478  GAAGCATATGTTTGTCCCATATTACAGCGCTGACATGCTAAATTCAGAGCAACAA 537

```

```

Db      8686  GAACATATGGCATTCCCATATTACACGCCATGATATGTCGCAAAATCCCTGAAACAGAA  8627
Oy      538    AACATCTCTCAATTTAAAGTGCCTCAATTCATGATGAGCAATCAAAACATTTATCG  597
Db      8626  AAAATATGAAAAATATCAAGTTCCTGATTCGATTCGATCCACATATTAATATCTCTCT  8567
Oy      558    GCAAAAGGGTATGATTAAGTCAAGGCAACTTAATAGATTAGATGATGGAATAGTGGCCA  657
Db      8566  GCAAAAGG-----CTGAGCGTTTGGGACAGCTGGGCA  8534
Oy      658    CTGCAAAACGCTGATGCTACTCGGCAATTTATCATGATATCATGCTCTCCCTTTA  717
Db      8533  TTACAAAACGCTGAGCGAGCTGCGCAACATATCAAGGCTACACATGCTCTTTCATTA  8474
Oy      718    GCAGGTGACCCAAAACAGTATGATATCTCCACTTATTTATTTATTTATCAAAAAGTCGGT  777
Db      8473  GCGGAGATTCCTAAATAATGCGGATACACATCGATTTTACATGCTCTATCCAAAATCGCG  8414
Oy      778    GATCATGATGATGACAGCTGGAATAATGCTGGAAGATTTTGAAGATATGATTAATTT  837
Db      8413  GAAATCTTATTTAGACGTGGAATAATGCTGCGCGCTTTTAAAGACAGCAAAATTC  8354
Oy      838    GTTCCAATGATCCCTATCTTAATATATCAACACAGAGTGGTCAAGTTCTCTACTTTA  897
Db      8353  GATGCAATGATTTCTATCTTAAAGACCAACACAAAGATGGTCAAGTTTCAAGCATTT  8294
Oy      898    ACCAAGATGCGCAAGTCCGTTTATTTATACAGATTTACTGAGTATCTGAAATGATG  957
Db      8293  ACATCTGACGGAATAATCCGTTTATTTTACACTGATTTCTCGGTAACATTA-----  8241
Oy      958    GGAACCGGTGCTGTATCACAATCTTTCAACGCTCAAGTAACTTATCCAGCCGAT  1017
Db      8240  -----CGGCAAAACAACTGACAACTGACAACTGACAAAGTTAACGTATCACATCAGAC  8192
Oy      1018  GCAGTACACTTAAGTGGATGAGTATCTGATCTATCTATCTGTTGATGAGCGGAGAC  1077
Db      8191  AGCTCT---TTGAACATCAACGGTGTAGAGATTAAATCAATCTTTGA---CGGTGAC  8138
Oy      1078  GGTACAGTTTATCAAAATATTACAGCAATTTATCATGATGAGCAAGTATGATTCAGTGT  1137
Db      8137  GGAATAAGGTATCAAAATATGATGAGCAATTTATCATGATGAGCAAGTATGATTCAGTGT  8078
Oy      1138  AACCATCTTTAAGAGACCTCATATGTTGAAGTAAAGGCGCATTAATATCTTCTTT  1197
Db      8077  AACCATGCTGAGATCTCTACACTAGTAGAAGTAAAGGCGCAAAATATCTTATTT  8018
Oy      1198  GAAGGATCTACTGGAACAACAGATGTTATCAAGCGGATCACTTTTCAATATTAAGCT  1257
Db      8017  GAAGCAAACTGTGAAGTGAAGATGCTACCAAGGCGAAGATCTTTATTATTAACAAACA  7958
Oy      1258  TACTATGCGGAGAGTGAAGTCTTCTTCCAGATGAAAAAAATTAACCTGCTTCAAGTCT  1317
Db      7957  TACTATGCGGAGAGTGAAGTCTTCTTCCAGATGAAAAAAATTAACCTGCTTCAAGTCT  7898
Oy      1318  AAAAAACAATGCTTCTTTAGCGAATGTCATTTAGCATTTGTAATTTGCGCATGAC  1377
Db      7897  AAAAAACGCGGCTGAGTTAGCAAAACGCGGCTCGGATATGATGAGCTAAACATGAT  7838
Oy      1378  TATACAGTAAAGTGTATGAACCATTTAGTGCATCAACACAGTACGAGATGAAGTC  1437
Db      7837  TACACAGTAAAGTGTATGAACCATTTAGTGCATCAACACAGTACGAGATGAAGTC  7778
Oy      1438  GAAGCGCGCATATATTAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA  1497
Db      7777  GAAGCGCGCATATATTAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA  7718
Oy      7717  TCAGAAATGAGATGAGCGCATTTACGCTTAACGATTTTACATCTGTTATGTTTCT  7658
Db      1558  GATCCTTAATTTGGCCACACACACACACACACACACACACACACACACACACACAC  1617
Oy      7657  AATCTTTAACTGGCCATACAAAGCCGCTGAACAAAACTGGCTTGTAAATGAT  7598

```

```

Oy      1618  CTTGACCCCTGATCTCACACACACTTACTCTCATTTGCGGTATCCCGCACCTGAAGT  1677
Db      7597  CTTGATCTTAAGATGTAACCTTTACTTACTACACTTCCGCTGATCCTCAAGCAAGGA  7538
Oy      1678  AATAATGTGATCTCACAAAGTTATATGACGAATAGAGCTTCTATTCAGAACATCTCT  1737
Db      7537  AACATGTGCTGATTAACAAGCTATATGACAAACAGAGATTTCTACGACAAACATTA  7478
Oy      1738  CACTGCGCGGACACTTGGGTTAATATTAAGGCTGTGACACATCTGAGAGAAAT  1797
Db      7477  ACGTTTGGCGCGAGCTTCTGCTGAACATCAAAAGGCAAGAAACATCTGTTCAAGAC  7418
Oy      1798  ACTTCC---GCACAAGACAAATTTCCATAGCATTAATCCCAATTAAGAAATGTC  1852
Db      7417  AGCATCTTGAACAAAGACATTAACATGTTAACAATTAATAAAGCAAAAGAAATGCC  7360

```

## RESULT 4

AA257329  
ID AA257329 standard; DNA; 1668 BP.

AA257329;

03-APR-2000 (first entry)

SacB and cpy chimeric gene #1.

Levansucrase; SacB; cpy; salt tolerance; vacuole guide peptide;

carboxypeptidase A; chimeric gene; transgenic plant; yeast;

Bacillus subtilis 168; Saccharomyces cerevisiae X8;

drought resistance; ds.

Chimeric - Saccharomyces cerevisiae.

Chimeric - Bacillus subtilis.

CN1231337-A.

13-OCT-1999.

08-APR-1998; 98CN-0101336.

08-APR-1998; 98CN-0101336.

(GENE-) INST GENETICS CHINESE ACAD SCI.

Chen S, Zhang H, Dong W;

WPI; 2000-087902/08.

Method for improving plant salt resistance using transgenic technology

Claim 6; Page 2; 12pp; Chinese.

A method has been developed for raising salt tolerance of plants using transgenic technology. The method comprises: cloning levansucrase gene SacB and vacuole guide peptide (cpy) sequence of carboxypeptidase A from Bacillus and DNA of yeast, using the two kinds of gene to create a chimeric gene, and using the chimeric gene to structure plant expression carrier, using the obtained plant expression carrier to transform a plant and screen for resistant seedlings. The method can obtain drought-resistant salt-tolerance plants, and the polymerase chain reaction (PCR) and Northern analysis of transgenic plants shows that the exogenous gene is integrated in the salt-tolerance transgenic plant. The present sequence represents a SacB and cpy chimeric gene from the present invention.

Sequence 1668 BP; 578 A; 372 C; 325 G; 393 T; 0 other;

Query Match 24.0%; Score 577.8; DB 21; Length 1668;

Best Local Similarity 66.4%; Pred. No. 5.2e-153; Matches 916; Conservative 0; Mismatches 412; Indels 51; Gaps 4;





Matches	916; Conservative	0; Mismatches	412; Indels	51; Gaps	4;
Qy	421 GTAGAGCGGAGCGACCCATATTTTTCGCGCAAAATGAAACAGCGGAGCTACAGGAA				480
Db	335 GCAAGTGAACCTATACAGCTTCGTCTCAATTAAGAAAGCAAAAGCCATATATGAA				394
Qy	481 GACTATGGTTTGGCCATATTTATACAGCGGCTGACATGCTTAAATTCAGAACAAAC				540
Db	395 ACATATCGCATTTCCATATTTACAGCGCATGATATGCTGCAAAATCCCTGAAACAGAA				454
Qy	541 AGTCTCAATTTAAAGTCCCTCATATTCATGATCAGCAATCAAAACATGATTGGCA				600
Db	455 AATGAAAAATATGCAATTCCTGAAATTCGATTCACAAATTAATAATATCTCTCTCA				514
Qy	601 AAGGGTATGATTAAGTACGCACTTAATATGATTTAGATGATGAGATAGCGGCACTG				660
Db	515 AAAAG-----CTGGACGTTTGGGACAGCTGGCCATTA				547
Qy	661 CAAAACGCTGATGCTACTGCGCAAAATATGATATGATATCAGATGCTCCGCTTTACA				720
Db	548 CAAAACGCTGACGCACTGTCGCAAACTATACAGGCTCCACATGCTCTTGCAATTAGCC				607
Qy	721 GGTATCCCAAAAACAGTATGATATCTCACTTATTTCTATCAAAAAGTCGCTGAT				780
Db	608 GGAATATCTTAATAATGCGGATGACACATGATTTACATGTTCTATCAAAAAGTCGCGCA				667
Qy	781 ACATGATTTGACGCTGGAATAATGCTGAGAGATATTGAAGATATGATATGATATTTGTT				840
Db	668 ACTTCTATTTGACGCTGGAATAATGCTGAGAGATATTGAAGATATGATATTTGTT				727
Qy	841 CCAATATGATCCGTATCTTAATATCAACACAGAGATGTCAGTCTGCTACTTTAAC				900
Db	728 GCAATATGATTTCTATCTTAATAAGCAACACAGAGATGTCAGTCTGCTACTTTACA				787
Qy	901 AAAGATGGCCAAAGTCGTTTATTTCTATGATATCTACTAGTATCCCGAAGATGTTGA				960
Db	788 TCTGACGGAATAATCCGTTTATTTCTACACTGATTTCTCGGTAACATTA-----				837
Qy	961 ACCGCTGCTGTAACCAATATCTTCACTGCTCAAGTAATATCCACCGGATGCA				1020
Db	838 -----CGGCAAAACAAACACTGACACTGCAACCTGCAATTAACATGACATCAGACGC				889
Qy	1021 GCTACACTTAAAGTCGATGAGTATCTGATCATTAATCTGCTTGTGATGCGGACGCT				1080
Db	890 TCT---TTGAACATCAACGCTGATGAGATTAATCAATCTTTGA---CGGTACGCA				943
Qy	1081 ACAGTTTATCAAAATATATACCAATTTATGATGATGAAGCAAGTGTTCAGGATTAAC				1140
Db	944 AAAACGTTTACAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG				1003
Qy	1141 CATACCTTAAAGAGACCTCACTATGTTGAAGATGAAGGCACTAAATATCTGCTTTGAA				1200
Db	1004 CATACGCTGAGAGATGCTCATACGTAGTAAGATTAAGGCACTAAATATCTGCTTTGAA				1063
Qy	1201 GCGAATCTGGAACACAGATGTTATCAGCGATCAGTCTTTCATTAATTAAGCTTAC				1260
Db	1064 GCAAACTGTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG				1123
Qy	1261 TATGCGGAAGTGAAGCTCTTCTTCAGAAATGAATAAATAAGCTTCAAGTCTTAA				1320
Db	1124 TATGCGGAAGTGAAGCTCTTCTTCAGAAATGAATAAATAAGCTTCAAGTCTTAA				1183
Qy	1321 AAMCAATTTGCTTTTAAAGCAATGATGATGATGATGATGATGATGATGATGATGATG				1380
Db	1184 AAMCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG				1243
Qy	1381 ACATGAAAAAGTGTATGAACCTTATGCTCATCAACACAGATGATGATGATGATGATG				1440
Db	1244 ACATGAAAAAGTGTATGAACCTTATGCTCATCAACACAGATGATGATGATGATGATG				1303
Qy	1441 CCGCCCAATATATTTAAATGAATATTAATGATGATGATGATGATGATGATGATGATG				1500
Db	1304 CCGCCCAATATATTTAAATGAATGAATGATGATGATGATGATGATGATGATGATGATG				1363

Qy	1501 AAAATGACGATGATGATTAACGCAAAAGATGTTATATGATGAGCGCCGAGCGAC	1560
Db	1364 AAAATGACGATGATGATGATTAACGCAAAAGATGTTATATGATGAGCGCGGAGCGAC	1423
Qy	1561 TCCTTAAATGGCCACACACACCCGATTAATGAATGATGATGATGATGATGATGATGATG	1620
Db	1424 TCCTTAAATGGCCACACACACCCGATTAATGAATGATGATGATGATGATGATGATGATG	1483
Qy	1621 GACCTGCTGATCTTACACACACTTCTCATTTGCGGTATCCGACCCGGAAGTAT	1680
Db	1484 GATCTTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1543
Qy	1681 AATGCTGATCTTACACACACTTCTCATTTGCGGTATCCGACCCGGAAGTAT	1740
Db	1544 AATGCTGATCTTACACACACTTCTCATTTGCGGTATCCGACCCGGAAGTAT	1603
Qy	1741 CTGCGGACAAAGCTTTGGGTTAATTAATGAAGGCTGACACATCTGAGAGAAATAG	1799
Db	1604 TTTGCGCAACGCTTCTCTGATCAATCAAGGCAAGAAACATCTGTTCAAGACAG	1662

RESULT 6  
 AAS68687  
 ID AAS68687 standard; cDNA; 2568 BP.  
 XX  
 AC AAS68687;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #4491.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 XX  
 PT P-PSDB; ABG04500.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 1: SEQ ID No 4491: 103pp: English.  
 CC  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPD  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 2568 BP: 827 A; 582 C; 615 G; 544 T; 0 other:

Query Match 23.4%; Score 564.2; DB 23; Length 2568;  
 Best Local Similarity 64.7%; Pred. No. 4.7e-149;  
 Matches 962; Conservative 0; Mismatches 468; Indels 57; Gaps 6;

```

OY 357 GACGATGACCTTAAAGATGGCGAAAAAAGCAGCTCCCTTACCTCAGAGTGTAT 416
DB 3 GAACATGAACATCAAAAAATTTGTAACAGCCAGCTTCTGACTTTTACGACTGCAC 62
OY 417 ATTAGTAGAGCGGAGCGACGCAATATTTTGGCGAGAAATGAACAGCGGGACTACA 476
DB 63 TCTGGCAGGA---GGAGCGACTCAAGCCCTCGCGAAAGAAATTAACAAAAAGCATACA 119
OY 477 GGAACATATGTTTGGCCATATTACAGCGGCTACATGCTAAAAATTTCCAGACACA 536
DB 120 AGAAACGTAAGCGGCTCTCTCATATTACAGCCCATATATGCTGAGATCCCTAAACACA 179
OY 537 AAACGTCCTCAATTTAAAGTGCCTCAATTCATCATGACGACCAAAACATGATATTC 596
DB 180 GCAAAACGAAAAATACCAAGTCCCTCATTCATCATCAACGATTTAAATAATTTGAGTC 239
OY 597 GCGAAAGGGATGATAGTACGACACTTAATAGATTAGTATGAGGATAGTGGCC 656
DB 240 TGCAAAAG-----ACCTGATGTGTGGAGACAGCTGGCC 272
OY 657 ACTGAAACGCGTATGCTACTCGGCAAAATATCATGATATCATACGTCCTCCGCTT 716
DB 273 GCTGAAACGCGTATGCTACTCGGCAAAATATCATGATATCATACGTCCTCCGCTT 332
OY 717 AGCAGTGACCCAAAAAAGCAGTATGATATCTCAATTCATTTCTATCAAAAAAGTCGG 776
DB 333 TGGGGAGGCCGGAAGAGCGTATGACATCATCATATGTTTATCAAAAGTGG 392
OY 777 TGATATCATGATGACAGCTGGAAAAATGCTGGAAGATATTGAATATGATAAAT 836
DB 393 CGACATCTCAATCGACAGCTGGAAAAAGCGGGCGCTGTTTAAAGACAGCATAGTT 452
OY 837 TGTTCGAATGATCGCTATCTTAATATCAACAGAGAGTGGTACAGTCTGACTT 896
DB 453 CGAGCGCAACGATCGATCTGAAAGATCAGACGAAAGATGTCGGTCTGCAACCTT 512
OY 897 AACCAAGATGCCAAGTCCGTTTATCTATACAGATTACTAGTAAATCCGAGATAG 956
DB 513 TACATCTAGCGAAAAATCCGTTTATCTACACTACATTCGGTAAACATTA----- 566
OY 957 TGGAAACGGTGTGTGTAACCAATCATTTCAACTGCTCAAGTAAATCTCCAGCCGA 1016
DB 567 -----CGGCAAAACAAAGCCCTGACAAACAGCAGAGTAATGTCTAAATCTGA 614
OY 1017 TGCAGCTACCTTAAGTCGATGAGTATCTGATATTAATGCTTGTGATGGGGAGA 1076
DB 615 T-----GACACACTCAAAATCAAGGAGTGAGATACAAAAGATTTTGA---CGGAGA 668
OY 1077 CGGTACACTTTATCAAAATATTTACAGCAATTTATCGATAGAGCAAGTATTCAGTGA 1136
DB 669 CGGAAAAACATATATGAGAGTTCACAGCTTATATGATGAGGCAATATATACCTCGGGA 728
OY 1137 TAACCATCTTTAGAGACCTCACTATGTTGAGATAGAGGCGCAATAATATCTTGTCT 1196
DB 729 CAACCATAGCGTGAAGACCTCACTACGTGAAACAAAGCCCATTAATATCTGTATTT 788
OY 1197 TGAACGGAATCTGAAACAGATGTTATCAAGGCAATCGTCTTCAATATAAAGC 1256
  
```

```

DB 789 CGAAGCCACACAGGGAAACGAAAAAGATTAACCAAGCGGAGAAATCTTTATTAACAAAGC 848
OY 1257 TTACTATGGCGGAGAGTACGCTTCTTCCACATGAAAAAATAAAGCTGTTCAAGTCC 1316
DB 849 GTACTACGCGCGCGGACGAACTCTTCCGTAAAGAAAAAGCAGAAAGCTTACGACGAGCC 908
OY 1317 TAAAAACAAATTCCTTTAGCGAATGTGATTTAGCAATTTGTAATGGCCGATGA 1376
DB 909 TAAAAAACGCGATGCTGAGTTAGCGAAGCGGCCCTCGGTATCATAGAGTTAATAATGA 968
OY 1377 CTATACAGTAAAGTGTATGAAGCCATGTAGTGCATCAACACAGTACAGATGAAGT 1436
DB 969 TTACACATTTGAAAAAGTAATGAAGCCGCTGATCACTTCAACAGGTAAGTATGAAT 1028
OY 1437 CGAAGCGCCCAATATATTTAAATGAATAATATGATATCTATTCACGATTCAGAG 1496
DB 1029 CGAGCGCGCGAATGTTTCAAAATGAAGCGCAAAATGTACTGTTCACTGATTCACGGG 1088
OY 1497 ATCCAAATGACAGATGTAATTAACGACAAAGATGTTATATGCTAGGGCCCGAGG 1556
DB 1089 TTCAAAATGACAGATGATGATTAACCAAGCATATTTACATGCTTGTATGTATC 1148
OY 1557 CGACTCCTTAATGCGCCACACACACACCCGATTAATGAATGAACTGACTTGTATGAACATA 1616
DB 1149 AAACCTTTAAACGCGCCTTACAAAGCGCTAACAACAAACAGGCGTGTGTCAAATGG 1208
OY 1617 TCTTGACCCCTGATCTCACACACACTTACTCTATTCGCGATATCCGCAACCTGAAG 1676
DB 1209 TCTTGATCAAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1268
OY 1677 TAATATGTGTACTCACAGATTTATGACGAAATGAGGCTTCTATCCAGAACATCAGT 1736
DB 1269 CAACATATGTTTATCAACACACTACATGACAAACAGAGGCTTCTGAGGATTAATAAGGC 1328
OY 1737 TCACCTGGGGGACAGCTTGGGGTAAATATTAAGGGCTGACACATCTGAGAGGAGAAA 1796
DB 1329 AACATTTGCCGCAACCTTCTTAATGACATCAAGGCAATAAACATCCGTTGTCAAAAA 1388
OY 1797 TAGTTC---GGACAAAGACAATTCCTATAGCGATATCTCCCAATA 1840
DB 1389 CAGCATCTGAGAGCAGACAGCAGCTGACATCATATGATGCTCCAGTA 1435

RESULT 7
AAS68494
ID AAS68494 standard; cDNA; 1584 BP.
XX
AC AAS68494;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4298.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: ABG04307.
  
```

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1: SEQ ID No 4298; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 1584 BP: 539 A; 358 C; 350 G; 337 T; 0 other:

Query Match 23.4%; Score 563.4; DB 23; Length 1584;

Best Local Similarity 65.0%; Pred. No. 6.1e-149; Mismatches 451; Indels 54; Gaps 5;

Matches 938; Conservative 0; Mismatches 451; Indels 54; Gaps 5;

QY 357 GAGCATGAACCTCAAAAGATTGGCGAAAAAGACAGCTCCCTTAACCTTCAGGACTGCTAT 416  
DB 162 GACATGAAACATCAAAAAATTGTAACAGCCACATCTTGACTTTAGACGACGACT 221  
QY 417 ATTAGTAGAGCGAGCGACCGCATATTTTGGCCAGCAAAATGAACACGCGGACTACAA 476  
DB 222 TCTGGCAGGA---GGAGGACTCAAGCCCTTCGCGAAAAATAACCAAAAAAGCATACAA 278  
QY 477 GGAAGACATATGTTTGGCCATATTACACGCGCTGACATGCTAAAAATTCAGACACAA 536  
DB 279 AGAAGCTAGCGGCTCTCTCTATTACAGCCATGATATGCTGAGATCCCTTAACACCA 338  
QY 537 AAACAGTCCCTCAATTAAAGTGCCTCAATTCATGATGACATCAAAACATTTGATTC 596  
DB 339 GCAAAAGCAAAATATCCAAAGTGCCTCAATTCATGATGACATCAAAATATTTGATTC 398  
QY 597 GGCAAAAGGATATAGTAGTACGCACTTAATAGATTAGATGTATGGATAGCTGCC 656  
DB 399 TGCAAAAGG-----ACTTGATGTGTGGGACAGCTGCC 431  
QY 657 ACTGCAAAAGCGTATGCTAGCTGCGCAAAATTCATGATGATACATCGTCCGCTT 716  
DB 432 GCTGCAAAAGCGTATGCTAGCTGCGCAAAATTCATGATGATACATCGTTCGTTGCTCT 491  
QY 717 AGCAGGTGACCCAAAAACAGTATGATACCTTCATTATTCTATCAAAAAAGTCCG 776  
DB 492 TGCAGGAAGCCGGAAGAGCTGATGACATCATCTACATGTTTTCAAAAGGTGG 551  
QY 777 TGATACATCGATTACAGCTGGAAAAATGCTGGAAGATATTGTAAGATATGATTAAT 836  
DB 552 CGACAACTCATGACAGCTGGAACGCGGCGCTCTTTAAAGCACGATTAAGCT 611  
QY 837 TGTTCAAATGATCGATCTTAAATATCAACACAGAGAGTCCAGCTTGTGCTACTTT 896  
DB 612 CGAGCCCAAGATCGATCTGAAAGATCAGACGCAAGATGTCGGTTCGCAACCTT 671  
QY 897 AACCAAGATGGCAAGTCCGTTATTTCTATACAGATTACTAGGTAACTCTGAAGATGG 956

DB 672 TACATCTACGSAANAATCCGTTTATCTACACGACATATCCGTAACATTA----- 725  
QY 957 TGGAAACCGGTGCTGTAACCAATATCTCAATGCTCAGTAACATTAATCCAGCCGA 1016  
DB 726 -----CGCAAAACAAAGCCTGACAAAGCCGAGGATTAATGTCAAAATCTGA 773  
QY 1017 TGCAGCTACCTTAAGTGCATGAGTATGATCATTAATCTGCTTGTGATGGCGGGA 1076  
DB 774 T---GACACCTCAAAATCAACGAGTGAAGATCACAACAGATTTTGA---CGGAGA 827  
QY 1077 CGGTACCTTTTAAATTAATTCAGCAATTTATGATGAGGCAAGTGGATTTTCAGTGA 1136  
DB 828 CGGAAACCAATATCAAGACGTTCCACGACTTTATGATGATGAAGCAATTAATATCCGCGGA 887  
QY 1137 TACCATATCTTTAGAGACCTCTACTATGTTGAAGATAGGCGCATTAATATCTTGTCTT 1196  
DB 888 CAACCATACGCTGAGAGACCTCTACTAGTTGAAGACAAAGCCATTAATATCTTGTAT 947  
QY 1197 TGAAGCAATPCTGSAACACAGATGTTATCAAGGCGATCAGTCTTCATTAATTAAGC 1256  
DB 948 CGAAGCCCAACAGGSAACAGAAAGGATACCAAGCGAGATCTTTATTAACAAAGC 1007  
QY 1257 TTACTATGCGGAAGTACGCTCTTCCAGATGAAAAATAAACTCTTCAAAAGTCC 1316  
DB 1008 GTACTACGCGCGGCGACGAACTTCTCCGTAAGAAAGCCAGAACTTTCAGCAGAGCC 1067  
QY 1317 TAAAAAACAAATGCTTTTATAGCAATGTCGATTAAGCATTTGTTGATTTGCCGATGA 1376  
DB 1068 TAAAAAACGCAATGCTGATGATGAGCAAGCGGCGCTCGATCATAGAGTTAAATTAAG 1127  
QY 1377 CTATACAGTGAAGATGTTATGAACCATTTAGTGCATCAACACAGAGATGAAGT 1436  
DB 1128 TTACCATTTGAAAAAGTATGAAAGCCGCTGATCTTCAACACGGAATGATGAAGT 1187  
QY 1437 CGAAGCCCAATATATTTAAATGAATTAATATGATGATCTATTCACGAGTATGAAGG 1496  
DB 1188 CGAGGCGCGCATGTTTCAAAATGAAAGGCAAAATGATGTTCACTGATTAACGCGG 1247  
QY 1497 ATCCAAATGACAGATGATGATTAACGACAAAGATTTTATGCTAGGCGCCGAGG 1556  
DB 1248 TTTCAAAATGACAGATGATGATTAACCTCAACAGATTTTACATGCTTGTATGATTC 1307  
QY 1557 CGACTCCTTAATAGGCGCACCAACCGATTAATGAAGTGAAGTGTATTAAGCATGA 1616  
DB 1308 AAACCTTTTAAACGCGCTTACAAAGCCGCTGACAAACAGAGCTTGTGCTGCAATGG 1367  
QY 1617 TCTTGACCTGCTGATCTCACACACACTTACTCTATTCATTCACCGACCTGGAAGG 1676  
DB 1368 TCTTGATCAAAAGATGATGATGATCACTTACTCTACTCTCGAAGTGGCGCAAGCAAGG 1427  
QY 1677 TAATATGTTGTTACTCACAGATTAATGACGAATAGAGGCTTTATCCAGAAATCACTC 1736  
DB 1428 CAAACAATGTTGTTTCAACAACTACATGACAAACAGAGGCTTCTCGAAGTAAAGAGC 1487  
QY 1737 TCACCTGCGGGAACAAGTGTGGGTTAATTAATGAAGGCTGACACATCTGAGAGGAAA 1796  
DB 1488 AACATTTGCGCAAGCTTCTTAATGAACATCAAAAGCAATTAACATCTGTTGCAAAA 1547  
QY 1797 TAG 1799  
DB 1548 CAG 1550

RESULT 8  
AAS85834  
ID AAS85834 standard; cDNA; 2044 BP.  
XX AAS85834;  
AC 13-FEB-2002 (first entry)  
XX DNA encoding novel human diagnostic protein #21638.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 OS MO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001MO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 DR WPI: 2001-639362/73.  
 XX P-PSDB: ABG21647.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX Claim 1: SEQ ID NO 21638; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC atftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX Sequence 2044 BP; 673 A; 507 C; 438 G; 426 T; 0 other;  
 SQ

Query Match 23.4%; Score 563.4; DB 23; Length 2044;  
 Best Local Similarity 65.0%; Pred. No. 76-149;  
 Matches 938; Conservative 0; Mismatches 451; Indels 54; Gaps 5;

QY 357 GACGATGAACCTCAAAAGATTGGCGAAAGAACAGCTCCGTAACCTCAGAGCTGAT 416  
 DB 622 GAACGTGAACATCAAAAATTTGTAAGAACAGCCAGTTCGACTTTAGCAGCTGACT 681  
 QY 417 ATTAGTAGAGCGAGCGAGCGCATATTTTGGCGAGCAAAATGACAGCGGGAGCTACAA 476  
 DB 682 TCTGCGACGAGGAGCGAGCGCTTCGCGAAGAAATAATACCAAAAAGCATACAA 738  
 QY 477 GGAAGACATATGTTTGGCCATATTACAGCGGCTGACATGCTAAATAATCCAGGACACA 536  
 DB 739 AGAAACGTAAGCGGCTCTCATATTATTCAGCCATATATGCTGCGATCCCTAAACACCA 798  
 QY 537 AAACAGTCTCTAATTTAAGTGCCTCATATGATGATGATGATGATGATGATGATGATG 596  
 DB 799 GCAAAACGCAAAATATACAGTGCCTCATATGATGATGATGATGATGATGATGATGATG 858

QY 597 GCAGAAAGGATATAGTAAAGTACAGCACTTAATAGATTAGATGATGATGATGATGATGATG 656  
 DB 859 TGCAAAAGG-----ACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 891  
 QY 657 ACTGCAAAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 716  
 DB 892 GCTGCAAAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 951  
 QY 717 AGCAGTGAAGCCCAAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 776  
 DB 952 TGGGGAGAGCCCGAAGAGCGTGTATGATGATGATGATGATGATGATGATGATGATGATG 1011  
 QY 777 TGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 836  
 DB 1012 CGACAACTCAATCGACACTGGAAGAAACGCGGCGCTGCTTTAAAGACACAGCATTAAGTT 1071  
 QY 837 TGTTCGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 896  
 DB 1072 CGAGCCCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1131  
 QY 897 AACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 956  
 DB 1132 TACATCTACGGAAGAAATCCGTTATTTCTACATGATGATGATGATGATGATGATGATGAT 1185  
 QY 957 TGAACCGGCTGCTGTAACCAATCATTTCACTGCTCAAGTAAACCTTATCCAGCCGGA 1016  
 DB 1186 -----CGCAAAACAAAGGCTGACAAACGCGGAGGTAAATGATGATGATGATGATGATG 1233  
 QY 1017 TGCAGTCACTTAATAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1076  
 DB 1234 T--GACACACTCAAAATCAACGAGTGAAGATGATGATGATGATGATGATGATGATGAT 1287  
 QY 1077 CGGTACATTTATCAAAATATTTACAGCATTTATGATGATGATGATGATGATGATGATGATG 1136  
 DB 1288 CGGAAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1347  
 QY 1137 TAACCATCTTTAAGAGACCTCACTATGTTGAAGATAGAGGCCATTAATATCTTGTCTT 1196  
 DB 1348 CAACCATGCTGTAGAGACCTCACTATGTTGAAGATAGAGGCCATTAATATCTTGTCTT 1407  
 QY 1197 TGAAGCAATGCTGGAACAAACAGATGTTATCAAGGCGATGATGATGATGATGATGATGATG 1256  
 DB 1408 CGAAGCAACAGGGAACAGAAAGGATGATGATGATGATGATGATGATGATGATGATGATG 1467  
 QY 1257 TTATCATGCGAGAGTGAAGTCTTCTCCAGAAATGAAATTAATTAATTAATTAATTAATTA 1316  
 DB 1468 GTACTAGGCGGCGGACAGAACTTCTCCGTAAGAAAGCAGAAAGCTTACAGAGAGCGC 1527  
 QY 1317 TAAAAACAAATGCTCTTTAGCGAATGATGATGATGATGATGATGATGATGATGATGATG 1376  
 DB 1528 TAAAAACGCGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1587  
 QY 1377 CTATACATGAAAGTGTATGAAACCATTTAGTGCATCAACAGATGATGATGATGATGATG 1436  
 DB 1588 TTACATGATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1647  
 QY 1437 CGAAGCGCCCAATATTTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1496  
 DB 1648 CGAGGCGCGGATGTTTCAAAATGAAAGCAAGCAATGATGATGATGATGATGATGATGATG 1707  
 QY 1497 ATCCAAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1556  
 DB 1708 TTCAAAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1767  
 QY 1557 CGATCTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1616  
 DB 1768 ACATCTTTAAGCGGCGCTTACAGCGCTGTAACAAACAGGCGTGTGCTGCAAAATGGG 1827  
 QY 1617 TCTTGACCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1676  
 DB 1828 TCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1887  
 QY 1677 TAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1736

Db 1888 CAAAGATGTGGTATCACAAGCTACATGACAAACAGAGCGTCTTCGAGCAATAAAGGC 1947  
QY 1737 TCACCTGGGGGACAAAGCTTGGGGTTAATTTAAAGGCTCTGACACATCTGAGGAGAAA 1796  
Db 1948 AACATTGCGGCAAGCTTCTTAATGAAACATCAAAAGCAATAAACATCTCTGTCAAAA 2007  
QY 1797 TAG 1799  
Db 2008 CAG 2010

RESULT 9  
AAS87384/C  
ID AAS87384 standard; cDNA: 2044 BP.  
AC AAS87384;  
XX 13-FEB-2002 (first entry)  
XX DE DNA encoding novel human diagnostic protein #23188.  
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
OS Homo sapiens.  
XX WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI: 2001-639362/73.  
XX P-PSDB: ABG23197.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
PS Claim 1; SEQ ID NO 23188; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (I) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pcf\_sequences.  
SQ Sequence 2044 BP; 426 A; 438 C; 506 G; 674 T; 0 other;

Query Match 23.4%; Score 563.4; DB 23; Length 2044;  
Best Local Similarity 65.0%; Pred. No. 7e-149;  
Matches 938; Conservative 0; Mismatches 451; Indels 54; Gaps 5;  
QY 357 GACGATGAACCTTCAAAAGATTGGCGAAAAAGACAGTCCCTAACCTTCAGAGCTGTAT 416  
Db 1423 GAACATGAACATCAAAAAAATTGTAAACAGCCACATCTTGACTTTAGACTGCACCT 1364  
QY 417 ATTACTAGAGCGGACGACCGCATATTTTGGCAGCAATGAACAGCGGAGCTACAA 476  
Db 1363 TCTGCGACAGA--GGAGGACCTCAAGCCCTTCGCAAAAATAACCAAAAAAGCATACA 1307  
QY 477 GGAAGACTATGTTTGGCCATATTCACAGCGGCTGACATGCTAAATAATTCAGAGCAACA 536  
Db 1306 AGAAACGTACGGCGCTCTCATATTACAGCGCATATATGCTGCGAGATCCCTAAACACA 1247  
QY 537 AAACAGTCCCTCAATTTAAAGTGCCCTCAATTCATGCAATCAGCAATCAAAAACATTGATTC 596  
Db 1246 GCAAAAGCAAAAATATCCAAAGTGCCCTCATTTGATCAATCAACGATTAATAATTTGATGC 1187  
QY 597 GCGAAAAGGGTATGATAAGTCAGGCACTTAATGATTTAGATGTATGGATAGCTGGCC 656  
Db 1186 TCGAAAAAGG-----ACCTGATGTGTGGGACAGCTGGCC 1154  
QY 657 ACTGCAAAAGCGTGTAGTACGGGCAAAATTATCATGATGATCATCGTCCGCTTT 716  
Db 1153 GCTGCAAAAGCGTGTAGTACGGGCAAAAGTGTAGTATCAAGGCTATCAGTTGTGTTGCTCT 1094  
QY 717 AGCAGTGACCCCAAAAACAGTATGATACCTCACTTATTTATTTATTCAAAAGTCCG 776  
Db 1093 TCGGGGAAGCCCGAAAGAGCGGTGATGACACATCATCATGTTTATTCAAAAGTCCG 1034  
QY 777 TGATCATCATGATGACAGCTGGAAAAATGCTGGAAAGATTTTGAAGATATGATTAATT 836  
Db 1033 CGACAACTCAATCGACAGCTGGAAAAAGCGGCGCTGTTTAAAGACAGAGATTAAGTT 974  
QY 837 TGTCCAAATGATCGGTATCTTAATATCAAAACAGAGAGTGGAGGTGGTCTGACTTT 896  
Db 973 CGAGCGCAACGATCCGATCTGAAATGACAGCAAGATGTCGGTTCGCAACCTT 914  
QY 897 AACCAAGATGCGCAGTCCGTTTATCTATACAGATTACTCAGGTAATCCGTAAGATGG 956  
Db 913 TACATCTAGCGGAAAAATCCGTTTATCTACGCTACATTTCCGTAACATTA----- 860  
QY 957 TGGAAACGGTGTGTTAACCAGATCATTTCAACTCTCAAGTAACCTTAATCCACCGCA 1016  
Db 859 -----CGGCAAAACAAAGCCTGACAAACAGCGAGTAATGTGTCAAAATCTGA 812  
QY 1017 TGCAGTACACTTAAGTCGATGAGATATCGATCATTAATCTGCTTGAATGGCGGAGA 1076  
Db 811 T---GACACACTCAAAATCAACGAGTGAAGATCAAAAACGATTTTGA---CGGAGA 758  
QY 1077 CGGTACAGTTTATCAAAATATTCAGCAATTTATTCATGAAGCAAGTGGATTTTCAGTGA 1136  
Db 757 CGGAAACATATTCGAACGTTTACAGATTTTATGATGAAGCAATTTATACATCCGCGGA 698  
QY 1137 TAACCATCTTAAAGAACCTCTACTATGTGAAGATTAAGGCCATTAATATCTTGTCTT 1196  
Db 697 CAACCATACGCTGAGAGACCCCTACTAGCTTGAAACAAAGCCATTAATACCTTGTATTT 638  
QY 1197 TGAACGAAATCTGGAACAACAGATGTTTATCAAGCCATGCTCTTCATTAATAAAGC 1256  
Db 637 CGAAGCCAAACAGGGAAGAAAAGGATACCAAGGCAAGATCTTTATTTAATAAAGC 578  
QY 1257 TTACTATGCGGGAATGACAGTCTTCTTCAGATGAATAAAATAAATTAAGCTTCAAGTCC 1316  
Db 577 GTACTACGCGCGGCGACGAGACTTCTTCGTTAAGAAAAGCAGAAAGCTTACAGAGAGCGC 518  
QY 1317 TAAAAACAAATTTGCTCTTTAAGCAATGTGATTTAGCATTTGTTGAATTTGGCCGATGA 1376  
Db 517 TAAAAAAGCGATGCTGAGATTAGCGAAGCGGCCCTCGGTATCATAGATTAAATAATGA 458

```

OY 1377 CTATACAGTGAAGAAGTGTATGAAACCATTAAGTCGCATCAACACAGTAGCAGATGAGT 1436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 457 TTACATCATGAAAGTAATGTAAGCCGCTGATCACTTCAACACGCTAAGCATGATAAT 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1437 CGAAGCCGCAATATATTTAAATGATATATATGATATCTATTCAGGATTCAGAGG 1496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 397 CGAGCCGCGGAATGTTTCAAAATGAAACGCAAAATGCTTGTCTGATGATTCAGCGG 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1497 ATCCAAATGACGAGTGAATGATTAAGCAACAAAGTGTATATGCTAGGCGCCGAGG 1556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 337 TTCAAAATGACGATGATGATGATTAACCAACGATATTACATCTTGTTATGATATC 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1557 CGACTCTTAAATGCGCCACACACACCGATTAATGAATGAACTGATGTAATGAATGAA 1616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 277 AACCTTTTAAACCGCCCTTACAAAGCCGCTGAACAAACAGGCGCTGTCTGCAATGCG 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1617 TCTTGACCTGCTGATCTCACACACTTACTCTCTCTGCGATGCCGACCCCTGAAGG 1676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 217 TCTTGATCCAAAGATGATGATCACTTACTCTCTCTGCGATGCCGACCCGACCAAGG 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1677 TATATATGTGTAAGTCAAGTATATATGAGATAGAGCTTCTATCCAGAACATCACTC 1736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 157 CAACATATGTGTTATACAAAGCTACATGACAAACAGGCTTCTTGAGAGATAAAGGCG 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1737 TCACCTGCGGACACAGCTTGGGTTAATATTAAGGCTTGACACATCTGAGAGAGAA 1796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97 AACATTGGCCAGAGCTTCTTATGATCAATCAAGGCAATTAACATCCGTTGTCANAAA 38
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1797 TAG 1799
    ||
DB 37 CAG 35

```

## RESULT 10

```

ID AAS69777 standard: cDNA; 2325 BP.
XX AAS69777:
AC AAS69777:
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #5581.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: AB605590.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 5581; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

```

```

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnosis as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2325 BP; 713 A; 593 C; 566 G; 453 T; 0 other;
Query Match 23.4%; Score 563.4; DB 23; Length 2325;
Best Local Similarity 65.0%; Pred. No. 7,5e-149;
Matches 938; Conservative 0; Mismatches 451; Indels 54; Gaps 5;
OY 357 GACGATGACCTCAAAAGATGCGCAAAAAGCAGCTGCGGTAACCTTCAGAGCTGAT 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3 GACATGAAACATCAAAAATTTGTAAACACCCACAGTGTGACTTTAGACATGCACT 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 417 ATTAGTAGAGCGGACGACCGCATATTTTGGCGACGAAATGAACAGCGGAGCTACAA 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 63 TCTGCACGAG---GGAGCGACCTCAAGCCCTCGCAAAAGAAATTAACAAAAGCATACAA 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 477 GGAAGACTATGTTTGGCCATATTAACGCGGCTGACATGCTAAATTTCCAGGACACA 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 120 AGAAAGCTAGCGGCTCTCATATTTTACAGCGCATATTTGCTGAGATCCCTAAACAGA 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 537 AACAGTCCCTCAATTTAAAGTGCCTCAATTCATGCAATCAGCAATCAAAAACATTTGATT 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 180 GCAAAACGAAATATACCAAGTGCCTCAATTCATGCAATCAGCAATTAATAATTTAGAGTC 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 597 GCAAAAGGGTATGATTAAGTACGCACTTAATAGATTAGATGATGATGAGTACGTGCC 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 240 TCCAAAAGG-----ACTTGATGTGTGGAGACGCTGCGC 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 657 ACTGCAAAAGCTGATGTGACTGCGGCAATTTATCATGATATCATGCTCGCTTT 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 273 GCTGCAAAAGCTGTGACGAGACAGTAGCTGAATACAGGCTATACAGTTGTGTTGCTCT 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 717 AGCAGGTGACCCAAAACAGTAGATGATACCTGCTCATTTATTTCTAACAAGAGTGG 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 333 TGGGGGAAGCCGGAAGAGCGTGAATGACACATCATCATATGTTTATCAAAAGCTGG 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 777 TGATACATGATTTGACAGCTGGAAGAAATGCTGGAAGATATTTGAAGATATGATTAAT 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 393 CGACAACCTCAATCGACAGCTGGAAAAACGCGGCGCTGCTTTAAAGACAGCATAGTT 452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 837 TGTCCAAATGATTCGCTATCTTAATATCAACACAGAGTGGTGGAGTGTGCTACTTT 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 453 CGAGCGCAAGATCGATCTGTAAGATCAGAGCAAGATGCTGGGTTCTGCAACCTT 512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 897 AACCAAGATGCGCAAGTCCGTTTATTTATACAGATTACTCAGGTAATCTCGAAGATGG 956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 513 TACATCTAGCGAAATATCGTTTATTTACAGCTAGATATTCGGTAAACATTA----- 566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 957 TGGAAACGGTGTGTAACCAATCATTTCAACTGCTCAAGTAATATCCAGCCGGA 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 567 -----CGCAAAACAAAGCCTGACACACAGCGAGGTAATGTGTCAAAATCTGA 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1017 TGCAGCTACCTTAAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 615 T---GACACACTCAAAATCAACGAGTGAAGATCAACAAAGATTTTGA---CGAGGA 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 1077 CGGTACAGTTTATCAAAATATTTCAGCAATTATGATGACGAAGTGGATTTCAGTGA 1136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 669 CGGAAAACATATGAGAACGTTTCAGCAAGTTATGATGACGAAGTATATACATCCGGGA 728
QY 1137 TAACCATATCTTAAAGACACCTCACTATGTTGAAGATAGGCGCAATAATATCTTCTT 1196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 729 CACCAATACGCTGAGACCTCACTAGTTGAAGACAAAGCCATTAATATCTTGTATAT 788
QY 1197 TGAAGCGAATACGTGACCAACAGATGGTATATCAAGGCGATGAGCTTCAATATAAGC 1256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 789 CGAAGCAACGCGGACGAGAACAGAAAGGATACCAAGACAGAACTTTATTTAAACAAAGC 848
QY 1257 TTACTATGGGGAAGTACGCTCTTCTTCAGAAATGAAAAAATTAAGCTTTCAAAGTCC 1316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 GTACTAGGGGGGGGCAAGACTCTTCTCCGTAAGAAAGCCAGAAAGCTTACGACAGAGCC 908
QY 1317 TAAAAAACAAATTCCTCTTCTTATGCGAATGTGCTATGAGCATGTTGATGATGCCGATGA 1376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 909 TAAAAAACGCGATCTGATGAGCAAGCGCGCCCTCGTATCATAGCTTAATATAATGA 968
QY 1377 CTATACAGTGAAGGTATATGAAACGATTAAGTGCATCAACACAGATGAGATGAAGT 1436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 969 TTACACATTTGAAAAAGTAAATGAAGCCGCTGATCATCTTCAACAGGTAAGTGAAT 1028
QY 1437 CGAAGCGGCGCAATATATTTAAATGAATATATATGATCTATTCACGATTCAGAGG 1496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1029 CGAGCGGCGGCAATGTTTCAAAATGAAGCGCAATGAGTCTTGTCACTGATTCACGCGG 1088
QY 1497 ATCCAAATGACGAGTGTGAATTAACGACAAAGATGTTATATGCTAGGCGCGGAGG 1556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1089 TTCAAAATGACGATGCTGATGATTAACCTCAACGATTTATACATGCTTGTATGTATC 1148
QY 1557 CGATCTCTTAATGAGCCACACACCGATTAATGAAGTGCATGATGATGAATCA 1616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1149 AATCTCTTAAACCGCCCTTACAAAGCCGCTGATCAACAAACAGGCTTGTGCAAAATGG 1208
QY 1617 TCTTACCCCTGCTATCTCAACACACCTTACTCTATGTCGATGCCGACACCTGAGAG 1676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1209 TCTTGAATCAAAAGATGATGATCACTTACTCTCACTTCCAGTGGCGGCAAGCAAAAG 1268
QY 1677 TAATATATGCTACTCAACAGTATATGACGAATAGAGGCTTCTATCCAGATCACTC 1736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1269 CAAACAAATGTGTTTACAAACGCTACATGACAAACAGAGGCTTCTCGAGATTAAGGCG 1328
QY 1737 TCACCTGCGGAGCAAGCTTGGGGTTAATTAAGGGCTGACACATCTGAGAGAGAAA 1796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1329 AACATTTGCCCAAGCTTCTTAATGACATCAAAAGGCAATTAACATCCGTTGCAAAA 1388
QY 1797 TAG 1799
    ||
Db 1389 CAG 1391

```

```

RESULT 11
AAS75941
ID AAS75941 standard; cDNA; 2325 BP.
XX
AC AAS75941;
XX
XX 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #11745.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200175067-A2.
XX
XX 11-OCT-2001.
XX

```

```

PF 30-MAR-2001; 2001W0-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HXSE-) HXSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR P-PSDB: ABG11754.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1: SEQ ID NO 11745; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
SQ Sequence 2325 BP; 713 A; 593 C; 566 G; 453 T; 0 other;

```

```

Query Match 23.4%; Score 563.4; DB 23; Length 2325;
Best Local Similarity 65.0%; Pred. No. 7,5e+149;
Matches 938; Conservative 0; Mismatches 451; Indels 54; Gaps 5;
QY 357 GACGATGACCTCAAAAGATTTGGGAAAAAAGACAGCTGCCATACCTTCAGAGCTGCTAT 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 GAACATGAAACATCAAAAATTTGTAANAACAGCCAGCTTGACTTTAGACTGCACT 62
QY 417 ATTAGTAGAGACGCGACGCGCATATTTTGGCCAGCAATGAACAGCGGGAGCTACAA 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 TCTGCGCAGCA--GGAGCGACTCAAGCCTTGGCAAAATAATACCAAAAAGATACAA 119
QY 477 GGAAGACTATGCTTTGGCCATATTATACAGCGGCTGACATGCTTAATAATCCAGGACAA 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 AGAAGCTAGCGGCTCTCTCATATTACAGCGCATGATATGCGAGATCCCTTAACAGCA 179
QY 537 AAACAGTCCATTAATTAAGTGCCTCAATTCATGATGACCAATCAAAAATTTGATTC 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 GCAAAACGAAATAATACCAAGTGCCTCAATTCGATCAATCAAGATTAATAATATTGAGTC 239
QY 597 GGCAAAAGGATATGATAGTCAAGCACTTAATGATTTAGATGATGATGATGATGATGATG 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 TGCAAAAAG-----ACTGATATGTTGGGACACTGCGC 272
QY 657 ACTGCAAAAAGCTGATGCTAGTGGCGCAATTTATCATGATGATGATGATGATGATGATG 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 GCTGCAAAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 332
QY 717 AGCAGGTGACCCAAAAAACAGTATGATATCTCACTTATTTATTCATCAAAAAGTCCG 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 333 TCGGGGAGCCCGAAGACGCTGATGACATCAATCTACATGTTTATCAAAAGTCGG 392
Oy 777 TGATCATCGATTGACAGCTGGAAAAATGCTGGAAGAGATTGGAAGATATGATAAAT 836
Db 393 CGACACATCATCGACAGCTGGAAAAACGGGGCGCTGCTTTAAAGACAGCATTAATT 452
Oy 837 TGTTCAAATGATCGGATCTTTAAATATCAACACAGAGTGGTCAGGTTCTGCTACTTT 896
Db 453 CGAGCCCAACGATCGATCTGGAAGATCAGACGCAAGATGTCGGGTCTGCAACCTT 512
Oy 897 AACCAAAATGGCCAAAGTCGTTTATCTATACAGATTACTAGATTAATCCGAAGATGG 956
Db 513 TACATCTCTGCAAAAAATCGTTTATCTACACATCTACATTCGCTTAAACATTA----- 566
Oy 957 TGAACCGGCTGCTGATCAACCAATCATCTTCAACTCTCAAGTAAGTAACTTATCCAGCCGA 1016
Db 567 -----CGGCAAAACAAAGCCTGACACACAGCGAGGTAATGTTGTCAAAATCTGA 614
Oy 1017 TGCAGCTACACTTAAAGTCGATGAGATCTGATCATTAATCTGCTTTGATGGCGGAGA 1076
Db 615 T---GACACACTCAAAATCAACGAGTGAAGATCAGAAAACGATTTTGA---CGGAGA 668
Oy 1077 CGGTACATTTATCAAAATATTCAGCATTTATCATGAGGCAAGTGATTCAGTGA 1136
Db 669 CGGAAAAACATATCAGAACGTTACAGAGTTATCATGAAAGCAATATACATCCGCGGA 728
Oy 1137 TAACCATCTTTAAGAGACCTCTACTATGTTGAAGTAAAGGCCAATTAATATCTGTCTT 1196
Db 729 CAACCATCGCTGAGAGACCTCTACTAGTTCAACAGGCAATTAATACCTGTATTT 788
Oy 1197 TGAACGGAATCTGACACAGATGTTATCAAGGCGATCCTTTCAATATATAAGC 1256
Db 789 CGAACCCAAACAGCGGAAACAGAAACGATACCAAGCAAGATCTTTATTATACAAAGC 848
Oy 1257 TTACTATGGCGGAAGTGAAGTCTCTTCCAGATTAATAAATAACGCTTCAAGAC 1316
Db 849 GTACTACGGCGGCGACGAACTCTCTCCGTTAAAGCCGAAAGCTTCACAGACAGCGC 908
Oy 1317 TAAAAAACAATATGCTTTAGCGAATGTCATTAAGCATTTGTTGAATGGCGGATGA 1376
Db 909 TAAAAAAGCGCATGCTGAGTTAGCGAAGCGGCCCTCGTATCATAGATTAATATGA 968
Oy 1377 CTATACAGTGAAGATGTTATGAACATTAAGTCGCATCAACACAGATGAGTAAGT 1436
Db 969 TTACACATTGAAAAAGTAAGAACCCGCTGATCACTTCAAAACAGGTAACGTGAAT 1028
Oy 1437 CGAAGCGGCCAATATATTAATAATGAATATGATCTATTCAGCGATTCAGAGG 1496
Db 1029 CGAGCGCGCGAATGTTTCAAAATGACGGCAAAATGATCTTGTTCACCTGATTCAGCGG 1088
Oy 1497 ATCCAAATGACAGATGATGATTAACGACAAAGATGTTTATATGAGCGCCGAGG 1556
Db 1089 TTCAAAATGAGATGATGATGATTAATCAACAGATTAATTAACGCTTGTATGATC 1148
Oy 1557 CGACTCCTTAATGCGCCACACACACGATTAATAACTGACCTGATTTGAACATGAA 1616
Db 1149 AAACCTTTAAACCGGCTTCAACAGCCGCTGAACAAACAGGCGCTGTGTCGAATGG 1208
Oy 1617 TCTTGACCTGCTGATCTACACACATTAACCTATGCGGTATCCGACCCCTGAGG 1676
Db 1209 TCTTGATCAAAACGATGATGATTAATCTACTACTCTGAGAGCGCAAGCCCAAGG 1268
Oy 1677 TAAATATGTTACTACAGATTAATGACGAATGAGGCTCTTATCAAGAACATCACTC 1736
Db 1269 CAACATGTGGTTATCACAAGTATGACAAACAGAGGCTTCTTGAGGATAAAAAGC 1328
Oy 1737 TCACCTGGGGCAAGCTTGGGGTTAATTAAGGCTCTGACACATCTGAGAGGAA 1796
Db 1329 AACATTTCGCCAGCTCTTAATGACATCAAAAGCAATTAACATCCGTTGTCAAAAA 1388
Oy 1797 TAG 1799
Db 1389 CAG 1391

```

```

RESULT 12
AAQ10498
ID AAQ10498 standard; DNA; 2351 BP.
XX
XX AAQ10498;
AC
XX
XX 17-APR-1991 (first entry)
DT
XX
DE B.amyloliquefaciens sacB[Bamp] fragment.
XX
XX levanusucrase; sucrose regulation; Bacillus; SacB[Bamp]; ss.
XX
XX Bacillus amyloliquefaciens.
OS
XX
FH Key Location/Qualifiers
FH sig_peptide 863..950
FH mat_peptide 951..2279
FH /tag= a
FH /product= levanusucrase
FH terminator 2291..2327
FH /tag= c
FH /note= "rho-independent"
FH misc_feature 706..764
FH /tag= d
FH /function= putative regulatory region
PN
XX WO9100913-A.
PD
XX 24-JAN-1991.
XX
XX 20-JUN-1990; 90WO-US03348.
PF
XX 07-JUL-1989; 89US-0376474.
PR
XX
XX (DUPO ) DU PONT DE NEMOURS CO.
PA
XX
XX Nagarajan V, Tang LB;
XX WPI: 1991-051337/07.
XX P-PSDB; AAR10671.
DR
XX
XX Sucrose regulatable expression vector which can replicate
XX derived from non-Bacillus subtilis Bacillus species for
XX expression in many bacterial species
PT
XX
XX Example: Fig 3; 44pp; English.
PS
XX
XX The levanusucrase-encoding sequence was isolated from Bacillus
XX amyloliquefaciens lambda ZAP phage plaques. After three successive
XX screening and re-cloning steps, (using probes based on the sequence
XX of B.subtilis SacB), all the plaques were found to hybridise to the
XX probes. Clones 2A and 2C containing the putative scab[Bamp] gene
XX sequences were converted to Bluescript to give plasmids pBE300 and
XX pBE301, respectively. DNA was isolated from the plasmids and
XX digested with EcoRI in preparation for a southern transfer. The
XX same scab[Bamp] probes were used in the southern hybridisation and
XX both hybridised only with a 1.5kb fragment. Sequence analysis of
XX this fragment showed it to be the scab[Bamp] gene. The sucrose
XX regulatory region can be incorporated into stable vectors to
XX control expression of heterologous genes in transformed B.subtilis.
XX See also AAQ10497-7.
SQ
XX Sequence 2351 BP; 739 A; 524 C; 506 G; 582 T; 0 other;
XX
XX Query Match 23.4%; Score 563.4; DB 12; Length 2351;
XX Best Local Similarity 65.0%; Pred. No. 7.5e-149;
XX Matches 938; Conservative 0; Mismatches 451; Indels 54; Gaps 5;
Oy 357 GACGATGACCTTCAAAACATTTGGCGAAAAAGACAGCTGCTTAACCTTCAGACTGCTAT 416
II IIIII IIIII I I IIII I I I I I I I I I I I I I I I I I I I I I

```



```

Db      860 GAACATGACATCAAAAAATTGTAAGAACCCAGCTCTGACTTTTACGACTGCACCT 919
Oy      417 ATTAGTAGAGCGGAGCGACCGCATATTTTGGCAGCAAAATGAGACAGCGGGACTACAA 476
Db      920 TCTGGCAGAGA---GGAGGCACTCAAGCCTTGGCCGAAAGAAAATATACCAAAAAAGCTACAA 976
Oy      477 GGAAGCATGTGTTTGGCCATATTACAGCGCGTGACATGCTTAAAAATTCAGAGCAACA 536
Db      977 AGAAACGTACGCGCTCTCATATTACAGCCATGATATGCTGAGATCCCTTAACACAGA 1036
Oy      537 AAACAGTCCATATTAAAGTGCCTCATTCATGCAATCAGCAATCAAAACATTGATTC 596
Db      1037 GCAAAACGAAAAATACCAATGCTCAATTCATCAATCAATCAATCAATCAATCAATCAAT 1096
Oy      597 GCAAAAGGATATGATAGTACAGCACTTATAGATTAGATGATGATGATGATGATGATG 656
Db      1097 TGCAAAAG-----ACCTGATGTGTGTGGTACAGCTGGCC 1129
Oy      657 ACTGCAAAACGCTGATGCTACTGCGGCAAAATATCATGATGATGATGATGATGATG 716
Db      1130 GCTGCAAAACGCTGAGGAGACAGTATGATGATGATGATGATGATGATGATGATGATG 1189
Oy      717 AGCAGGTACCCCAAAAAACATGATGATGATGATGATGATGATGATGATGATGATG 776
Db      1190 TGCGGGAAGCCCGAAGAGCGTATGATGATGATGATGATGATGATGATGATGATG 1249
Oy      777 TGATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 836
Db      1250 CGACACATCATTCACAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1309
Oy      837 TGTTCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 896
Db      1310 CGAGCCCAAGCATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGAT 1369
Oy      897 AACCAAGATGCGCAAGTCCCTTATTTCTATACAGATTTACTAGATTTCTGAAAGATG 956
Db      1370 TATCATCTAGCAAAAAATCCGTTTATTTACAGTATGATGATGATGATGATGATG 1423
Oy      957 TGAACCCGCTGCTGTAACCAAAATCATTTCAACTCTCAAGTAACTTATCCACCGCA 1016
Db      1424 -----CGGCAACAAAGCTGACAAACGCGAGGATGATGATGATGATGATG 1471
Oy      1017 TGCAGTCACTTAAGTGCATGATGATGATGATGATGATGATGATGATGATGATG 1076
Db      1472 T---GACACACTCAAAATCAACGAGTGAAGATCACAAACGATTTTGA---CGGAGA 1525
Oy      1077 CGGTACATTTATCAAAATATTCAGCAATTTATGATGATGATGATGATGATGATG 1136
Db      1526 CGGAAACATATTCAGAGCTTCAGAGTTTATGATGATGATGATGATGATGATG 1585
Oy      1137 TAACCATCTTTAAGAGACCTCACTATGTTGAAGTAAAGGCAATTAATCTGTCTT 1196
Db      1586 CAACCATGCGTGAAGACCTCACTATGTTGAAGTAAAGGCAATTAATCTGTCTT 1645
Oy      1197 TGAAGCGAATCTGACACACAGATGTTATCAAGCGCATGCTTTCAATATATAAG 1256
Db      1646 CGAAGCCAAACGCGGAAACGAGATACCAAGCAAGATCTTTATTAACAAGC 1705
Oy      1257 TTACTATGCGGAGTACGCTCTTCCAGATGAAATAAATAACGCTTCAAGTCC 1316
Db      1706 GTACTACGCGCGCGACGAACTCTTCCGTAAAGCCAGAAAGCCATGACGACAGCGC 1765
Oy      1317 TAAAAAACAAATGCTCTTTAGCAATGATGATGATGATGATGATGATGATGATG 1376
Db      1766 TAAAAAACGCGATGCTGAGTAAAGCGCGCCCTCGGTATCATAGATTAATATAAG 1825
Oy      1377 CTATACAGTGAAGAGTGTATGAACCATATGATGATGATGATGATGATGATGATG 1436
Db      1826 TTACACATTTGAAAAAGTAAAGAGCCGCTGATCACTTCAAAACGAGTAACTGATGA 1885
Oy      1437 CGAAGCGCGCAATATTTTAAATGAATTAATGAATGATGATGATGATGATGATG 1496
Db      1886 CGAGCGCGCAATTTTCAAAATGAACGCAAAATGATGATGATGATGATGATGATG 1945

```

```

Oy      1497 ATCCAAATGACAGTGAATGAATTAACAGCAAAAGATTTATATCTAGGCGCGAG 1556
Db      1946 TTCAAAATGAGAGTGAATGATTAATCAACAGATTTATACATCTGGTATGATTC 2005
Oy      1557 GCATCTCTTAATGAGCCACACACACCCGATTAATGAACCTGATGATGATGATG 1616
Db      2006 AAACCTTTAAACCGCCCTTACAAAGCCCTGAAACAAACAGGCGTGTGCTCAATGG 2065
Oy      1617 TCTTGACCTGCTGATCTCACACACATCTCTCATTTGCGGTATCCCGCACCTGAGG 1676
Db      2066 TCTTGATCCAAAGATGATGATGATGATGATGATGATGATGATGATGATG 2125
Oy      1677 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1736
Db      2126 CAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2185
Oy      1737 TCACCTGCGGACAAAGCTTGGGTTAATTAAGAGTCTGACACATCTGAGAGAA 1796
Db      2186 AACATTTGCGCCCAAGCTTCTTAAATGACATCAAGCAATTAACATCTCGTTCAAAA 2245
Oy      1797 TAG 1799
Db      2246 CAG 2248

RESULT 13
AAS69133
ID AAS69133 standard; cDNA; 2370 BP.
XX
XX AAS69133;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #4937.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX P-PSDB: ABG04946.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations.
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1: SEQ ID NO 4937; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting and
XX quantitating a polypeptide in tissue, as molecular weight markers and as

```

CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [http://wipo.int/pub/publ/published\\_pcl\\_sequences](http://wipo.int/pub/publ/published_pcl_sequences).

XX Sequence 2370 BP; 743 A; 535 C; 544 G; 548 T; 0 other;

Query Match 23.4%; Score 563.4; DB 23; Length 2370;  
Best Local Similarity 65.0%; Pred. No. 7.5e-149;  
Matches 938; Conservative 0; Mismatches 451; Indels 54; Gaps 5;

```
QY 357 GACGATGAACCTTCAAAAGATGGCGCAAAAGACAGCTGCCCTTACCTTCAGAGCTGCTAT 416
DB 948 GAAACATGAACATCAAAAATTTGTAACACACAGCTTCGACTTTAGCAGCTGCACCT 1007
QY 417 ATTAGTAGAGCGGAGCGAGCCGATATTTTGGCAGCAATGAACAGCGGAGCTACAA 476
DB 1008 TCTGGCAGGA---GGAGGCACTCAAGCCCTCCGAAAGAAATTAACCAAAAGCATACAA 1064
QY 477 GGAAGCATATGTTTGGCCATATTACAGCGCTGCATGCTAAATAATCCAGGACACA 536
DB 1065 AGAAACGTAAGCGGCTCTCATATTACAGCCCATGATGCTGCGAGATCCCTAAACAGCA 1124
QY 537 AAACAGTCTCTAATTAAAGTCCCTCAATTCATGCATCAGCAATCAAAACATTTGATTC 596
DB 1125 GCAAAACGAAATATCAAGTCCCTCAATTCATGCATCAACGATTAATAAATATTGAGTC 1184
QY 597 GCGAAAGGGTATGATTAAGTACGCACTTAATAGATTAGATGATGAGTACGCTGCC 656
DB 1185 TCGCAAAAGG-----ACTTGATGCTGAGGAGCGTGGCC 1217
QY 657 ACTGCAAAACGCTGATGATGCTGCGCAAAATATCATGATGATCATGCTCCGCTTT 716
DB 1218 GCTGCAAAACGCTGAGGAGCAAGTACGATCAACAGGCTATCAGCTGTGTTGCTCT 1277
QY 717 AGCAGGTGACCCAAAAAACAGATGATGATCTCCACTTCAATTTATCTATCAAAAGTCCG 776
DB 1278 TCGGGGAAAGCCGAAAGAGCGTGAACATCAATCTACATGTTTATCAAAAGGTCGG 1337
QY 777 TGATACATCGATTGACAGCTGGAATAATGCTGGAAGATTTGAAGATATGATTAATT 836
DB 1338 CGACACATCAATCGACAGCTGGAAAAACCGGCGGTGCTTTAAAGACAGGATTAAGTT 1397
QY 837 TGTTCCAATGATCGGTATCTTAATATCAACACAGAGATGCTGATGCTGCTACTTT 896
DB 1398 CGAGCCCAACGATCCGATCTCTAAAGATGACGCAAGATGCTCGGTTCTGCAACCTT 1457
QY 897 AACCAAGATGGCCAGTCCGTTTATCTATACAGATTACTAGGTAATCTCGAAGATGG 956
DB 1458 TACATCTGACGAAAAATCCGTTTATCTACACTGATCTCCGGTAACATTTA----- 1511
QY 957 TGGAACCGGTGCTGTTAACCAATATCTTCACTGCTCAAGTAACTTATCCAGCCGGA 1016
DB 1512 -----CGGCAAAACAAAGCTGCAACACAGCGAGTAAATGTGTCAAAATCTGA 1559
QY 1017 TGCAGCTACATTAAGTCGATGAGATCTGATCATTAATGCTTGTGATGGGGGAGA 1076
DB 1560 T---GACACACTCAAAATCAACGAGATGAGATCAACAAACGATTTTGA---CGGAGA 1613
QY 1077 CGTACAGTTTATCAAAATATTCACCAATTTATCGATGAAGCAAGTGGATTTCAAGTGA 1136
DB 1614 CGAAACAAATATCAAGAACGTTACAGAGTTTATCATGAAGCAATTTATACATCGCGGA 1673
QY 1137 TAACATACCTTTAAGAGACCTCCTATGTTTAAGATTAAGGCCAATATATCTGTCTT 1196
DB 1196
```

```
DB 1674 CAACCATACGCTGAGAGACCCCTCACTACCTGTAAGAACAAAGCCATTAATACCTTGATTT 1733
QY 1197 TGAACCGAATCTGTGAACACAGATGTTATCAAGCGGATGAGCTTTCATTAATAAGC 1256
DB 1734 GAAACCCAAACAGGAGAACGAAAGATACCAAGGCGAAGATCTTATTAATAAAGC 1793
QY 1257 TTACTATGCGGAGATGACGCTCTTCTTCAGAAATGAAAAATAAAGTCTTCAAGTCC 1316
DB 1794 GTACTACGCGGCGGACAGCACTTCTTCGTAAGAAAGCCAGAAAGCTTACAGAGCGC 1853
QY 1317 TAAAAACAAATGCTTCTTTAGCAATGTCATTTAGCATTTGTTGATTTAGCCGATGA 1376
DB 1854 TAAAAACCGCATGCTGATGATGACCAAGCGCGCCCTCGTATCATAGATTAATAAATGA 1913
QY 1377 CTATACAGTGAATAAGTATTAACCAATTAAGTATGATGATCAACACAGATGAGTAAAGT 1436
DB 1914 TTACATTTGAAAAAGATTAAGAACCCCTGATCTCTTCAACACGATTAAGTAAAT 1973
QY 1437 CGAACGCGCAATATATTTAAATGAATTAATGAATGATATCTATTCACGATTCAGAG 1496
DB 1974 CGAGCGCGGAGATGTTTCAAAATGACGCGCAATGATGATCTGATTCATGATTCAGCGG 2033
QY 1497 ATCCAAATGACGATGATGATTAACGACAAATGTTATATGCTAGAGCGCGGAGG 1556
DB 2034 TTCAAAATGACGATGATGATTAATCAACGATTTATACAGCTTGTATGATATC 2093
QY 1557 GCATCTCTTAATGAGCCACACACACCAATGAATGAATGATGATGATGATGATGATG 1616
DB 2094 AAACCTTTTAACCGGCTTACACACCGCTGACACAAACAGGCTGTGCTGCAATGAG 2153
QY 1617 TCTTGACCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1676
DB 2154 TCTTGATCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2213
QY 1677 TATATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1736
DB 2214 CAACATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2273
QY 1737 TCACCTGCGGAGCAAGCTTGGGTTAATTAATTAATTAATTAATTAATTAATTAATTA 1796
DB 2274 AACATTTGCGCAACCTTCTTATATGATCAACATTAATTAATTAATTAATTAATTA 2333
QY 1797 TAG 1799
DB 2334 CAG 2336

RESULT 14
AAS92594
ID AAS92594 standard; cDNA: 2370 BP.
XX
AC AAS92594;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28398.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
```





```
Db      2158  TCTTGATCCAAAGCATGTGACATTCACTCTCACTTGCGAGTGCCGCAAGCCAAAGG 2217
OY      1677  TAATAATGTGTACTCACAAGTTATATAGCGAATAGAGGCTTCTATCCAGACATCAGTC 1736
Db      2218  CAGCAATGTGTTATCACAAGCTACATGACAAACAGAGGCTTCTTGAGGATAAAAAGGC 2277
OY      1737  TCACCTGCGGACCAAGCTTGGGGTTAATATTAAAGGGTCTGACACATCTGAGAGAGAAA 1796
Db      2278  AACATTGGCGCAAGCTTCTTATATGAACATCAAGGCAATAAACATCCGTGTCAAAA 2337
OY      1797  TAG 1799
Db      2338  CAG 2340
```

Search completed: June 18, 2003, 03:10:34  
Job time : 537.636 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 01:50:51 ; Search time 6277.52 Seconds  
(without alignments)  
1163.571 Million cell updates/sec

Title: US-09-986-682B-5

Perfect score: 2408

Sequence: 1 CGGGCAAAATCTAGATTCC.....TACCAATPAAAAATGCTGATC 2408

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenBank:\*

1: gb-ba:\*

2: gb-htg:\*

3: gb-in:\*

4: gb-om:\*

5: gb-ov:\*

6: gb-pat:\*

7: gb-ph:\*

8: gb-pl:\*

9: gb-pr:\*

10: gb-ro:\*

11: gb-sts:\*

12: gb-sy:\*

13: gb-un:\*

14: gb-vi:\*

15: em-ba:\*

16: em-fun:\*

17: em-hum:\*

18: em-in:\*

19: em-mu:\*

20: em-om:\*

21: em-or:\*

22: em-ov:\*

23: em-pat:\*

24: em-ph:\*

25: em-pl:\*

26: em-ro:\*

27: em-sts:\*

28: em-un:\*

29: em-vi:\*

30: em-htg-hum:\*

31: em-htg-inv:\*

32: em-htg-other:\*

33: em-htg-mus:\*

34: em-htg-pln:\*

35: em-htg-rod:\*

36: em-htg-mam:\*

37: em-htg-vrt:\*

38: em-sy:\*

39: em-htgo-hum:\*

40: em-htgo-mus:\*

41: em-htgo-other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2408	100.0	2408	1	AB010272
2	2408	100.0	2408	6	AR078182
3	2408	100.0	2408	6	AR208488
4	2408	100.0	2408	6	E15383
5	1365	56.7	1365	6	AR078181
6	1365	56.7	1365	6	AR208487
7	1365	56.7	1365	6	E15382
8	756.6	31.4	11318	1	AE007686
9	627	26.0	1449	6	AX34229
10	611.8	25.4	2007	6	BSSACBG
11	611.8	25.4	5941	12	SYNPRLC
12	611.8	25.4	6347	12	SYNPRLA
13	611.8	25.4	6798	12	SYNPRLB
14	611.8	25.4	9297	12	AF048702
15	611.8	25.4	88063	1	BS294043
16	611.8	25.4	209510	1	BS080018
17	610.4	25.3	5831	12	AF047518
18	610.4	25.3	5842	12	AF004910
19	610.4	25.3	5846	12	XXU17500
20	610.4	25.3	6349	12	AF047519
21	607.2	25.2	10317	6	AR162197
22	594.6	24.7	1422	1	BSU34874
23	563.4	23.4	2350	1	BASACB
24	563.4	23.4	11612	12	CVU80929
25	563.4	23.4	16011	12	XXU09128
26	563.4	23.4	16297	12	CVU75991
27	563.4	23.4	18772	12	AF133437
28	563.4	23.4	19500	12	CVU75992
29	563.4	23.4	22530	12	AB020028
30	563.4	23.4	107872	2	AC013485
31	563.4	23.4	110000	2	AC058788_2
32	563.4	23.4	113316	2	AC007866
33	563.4	23.4	122526	2	AC012647
34	563.4	23.4	132928	2	AC091089
35	563.4	23.4	133901	2	AC008146
36	563.4	23.4	146349	2	AC007863
37	563.4	23.4	151443	2	AC007865
38	563.4	23.4	154413	2	AC007862
39	563.4	23.4	155204	2	AC007926
40	563.4	23.4	156128	2	AC113265
41	563.4	23.4	156709	2	AC013484
42	563.4	23.4	156754	2	AC008031
43	563.4	23.4	157774	2	AC013353
44	563.4	23.4	159717	2	AC008368
45	563.4	23.4	162249	2	AC009259

## ALIGNMENTS

RESULT 1

AB010272

LOCUS AB010272

DEFINITION Bacillus sp. gene for beta-fructofuranosidase, complete cds.

ACCESSION AB010272.1 GI:3399690

VERSION

KEYWORDS Beta-fructofuranosidase.

SOURCE

ORGANISM Bacillus sp.

REFERENCE 1 (sites)

AUTHORS Kurimoto,M

TITLE Cloning and sequencing of beta-fructofuranosidase gene from

JOURNAL Bacillus sp. V230

Unpublished

## REFERENCE 2 (bases 1 to 2408)

Tsukaki, K.  
Direct Submission  
Submitted (08-JAN-1998) Keiji Tsusaki, Hayashibara Biochemical  
Laboratories, Inc., Amase Institute; 7-7 Amase-mitami machi,  
Okayama, Okayama 700, Japan (E-mail: amasehepo.harenet.or.jp,  
Tel: 086-231-6731, Fax: 086-231-6738)

## FEATURES

source

1. 2408  
/organism="Bacillus sp."  
/strain="V230"  
/db\_xref="taxon:1409"  
/note="Isolated from soil"  
361.1824  
/codon\_start=1  
/transl\_table=11  
/product="Beta-fructofuranosidase"  
/protein\_id="BAA32083.1"  
/translation="MNEKRLAKKAATVETRAILVGDGPHLFAQOMNSGDYKEDYF  
AHITRADMLKIPGOONSPQPKYPOFNASIKIKDSAKYDGSNLIIDLVMSPELON  
ADGTAAHGHYHIVSALAGDPKNSDTPHLFYOKVCDTSDSKNNAKRVEDMKPV  
PNDFLYKOTQDMSSTLTKDGVRLEYYTDISGNPEDEGDTAGQITISIAOVNLSOP  
DAATLVGVSDHKSVEFDGSGTYQNIQQLDESKWISGDHNLRLDHYVEDKHKY  
LVFEANTGTSTYKPLVASNTVADEVERANIEFMNKKVLTDSRGSKMTSDINDKY  
ELADYTVKSVKPLVASNTVADEVERANIEFMNKKVLTDSRGSKMTSDINDKY  
YMLGGGSLNGPNPILNETGLVLMNMLDPADLTHTYSHCGIPHEGNNVLTSTMTN  
RGFYEHSHLRDKIGVNIKSDTSNGSENSGQGF"

## CDS

## BASE COUNT

777 a 492 c 496 g 643 t

## ORIGIN

Query Match 100.0%; Score 2408; DB 1; Length 2408;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGGAAATTAATTAATTCATTCAGTCTCCCACTTGGTGAAGAGAGCTTCG 60  
DB 1 CGGGGAAATTAATTAATTCATTCAGTCTCCCACTTGGTGAAGAGAGCTTCG 60  
QY 61 GACTCCCATGAGCTGTGCGTAAGAAAACAGCTTACATGAGCAATTAAGAGAT 120  
DB 61 GACTCCCATGAGCTGTGCGTAAGAAAACAGCTTACATGAGCAATTAAGAGAT 120  
QY 121 TTCAAAAGTCTGGGATTAAGGCCAGTCCACTTTGAATTAACCTTTGACCTGTGTG 180  
DB 121 TTCAAAAGTCTGGGATTAAGGCCAGTCCACTTTGAATTAACCTTTGACCTGTGTG 180  
QY 181 GGGGCTTTTGTGTTTATTTATTCACATGCAAGTGTGCTCCCTATATCAATTTAA 240  
DB 181 GGGGCTTTTGTGTTTATTTATTCACATGCAAGTGTGCTCCCTATATCAATTTAA 240  
QY 241 GACGAAATTTCAATCAATCCATGCCATCCCAATTAACGTCCTCTATACTTTTAA 300  
DB 241 GACGAAATTTCAATCAATCCATGCCATCCCAATTAACGTCCTCTATACTTTTAA 300  
QY 301 TTAATTAAGAAATATCAAGAGCTTCTTATCAATTAATTAATTAATTAATTAATTA 360  
DB 301 TTAATTAAGAAATATCAAGAGCTTCTTATCAATTAATTAATTAATTAATTAATTA 360  
QY 361 ATGAACCTCAAAAGATTTGGCGAAAAGACAGCTGCGTAACCTTGAAGAGCTGTATTA 420  
DB 361 ATGAACCTCAAAAGATTTGGCGAAAAGACAGCTGCGTAACCTTGAAGAGCTGTATTA 420  
QY 421 GTAGAGGCGAGAGCGCATATTTTGGCGAGAAATGAAGAGCGGGGACTACAGGAA 480  
DB 421 GTAGAGGCGAGAGCGCATATTTTGGCGAGAAATGAAGAGCGGGGACTACAGGAA 480  
QY 481 GACTATGTTTGGCCATATTAACAGCGCGTGCATGTATTAATAATTCAGAGCAACAAAC 540  
DB 481 GACTATGTTTGGCCATATTAACAGCGCGTGCATGTATTAATAATTCAGAGCAACAAAC 540  
QY 541 AGTCTCAATTTAAGTGCCTCAATTCATGATGAGCAATCAAAACATTTATTCGCA 600  
DB 541 AGTCTCAATTTAAGTGCCTCAATTCATGATGAGCAATCAAAACATTTATTCGCA 600

DB 541 AGTCTCAATTTAAGTGCCTCAATTCATGATGAGCAATCAAAACATTTATTCGCA 600  
QY 601 AAAGGTTATGAATTAATCAAGCACTTAATTAATTAATTAATTAATTAATTAATTA 660  
DB 601 AAAGGTTATGAATTAATCAAGCACTTAATTAATTAATTAATTAATTAATTAATTA 660  
QY 661 CAAACGCTGATGTACGCGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720  
DB 661 CAAACGCTGATGTACGCGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720  
QY 721 GGTGACCCCAAAACAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780  
DB 721 GGTGACCCCAAAACAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780  
QY 781 ACATGATTTGACAGCTGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840  
DB 781 ACATGATTTGACAGCTGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840  
QY 841 CCAATGATCCGTATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900  
DB 841 CCAATGATCCGTATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900  
QY 901 AAAGATGGCCAAAGTCCGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 960  
DB 901 AAAGATGGCCAAAGTCCGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 960  
QY 961 ACCGCTGCTGGTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020  
DB 961 ACCGCTGCTGGTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020  
QY 1021 GCTACACTTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080  
DB 1021 GCTACACTTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080  
QY 1081 ACAGTTTATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140  
DB 1081 ACAGTTTATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140  
QY 1141 CATACTTAAGAGACCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200  
DB 1141 CATACTTAAGAGACCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200  
QY 1201 GCGAATTAAGAGACCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1260  
DB 1201 GCGAATTAAGAGACCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1260  
QY 1261 TATGCGGGAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320  
DB 1261 TATGCGGGAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320  
QY 1321 AAACAAATTTGCTTTAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380  
DB 1321 AAACAAATTTGCTTTAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380  
QY 1381 ACAGTGAAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1440  
DB 1381 ACAGTGAAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1440  
QY 1441 CGCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500  
DB 1441 CGCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500  
QY 1501 AAATGAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1560  
DB 1501 AAATGAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1560  
QY 1561 TCCTTAATTTGCGCCAC 1620  
DB 1561 TCCTTAATTTGCGCCAC 1620  
QY 1621 GACCTGCTGATCTAC 1680  
DB 1621 GACCTGCTGATCTAC 1680



1681 AATGTGACTACACAAGTTATATGACGATAGAGCTTCTATCCAGAACATCCTCTCAC 1740  
1681 AATGTGACTACACAAGTTATATGACGATAGAGCTTCTATCCAGAACATCCTCTCAC 1740  
1741 CTGGGGGCAAGCTTGGGGTTATATTAAGGCTGTACACATCTGAGAGAAATAGT 1800  
1741 CTGGGGGCAAGCTTGGGGTTATATTAAGGCTGTACACATCTGAGAGAAATAGT 1800  
1801 TCCGACAGAGACATATCCATAGGATATCTCCATATAAAGAAATGACAGGCA 1860  
1801 TCCGACAGAGACATATCCATAGGATATCTCCATATAAAGAAATGACAGGCA 1860  
1861 GATCTTACCTTTTCCCAACCTTTTAAAAAAGAGTTTGTACCACTTTAGAA 1920  
1861 GATCTTACCTTTTCCCAACCTTTTAAAAAAGAGTTTGTACCACTTTAGAA 1920  
1921 GAAAAAAGATCTTGGCGCGGCGCAATTACCATTGCCAAGAGCTACCAATAAAA 1980  
1921 GAAAAAAGATCTTGGCGCGGCGCAATTACCATTGCCAAGAGCTACCAATAAAA 1980  
1981 AGAAATTTTCCCTTTTGGGGGGGAGTCTTTTCTATCTGGGGTGGGATGTG 2040  
1981 AGAAATTTTCCCTTTTGGGGGGGAGTCTTTTCTATCTGGGGTGGGATGTG 2040  
2041 GCCCCACAGGAGGACCTTTATTTATTTATAGCATGATCTGGGATTTGCTGGCTA 2100  
2041 GCCCCACAGGAGGACCTTTATTTATTTATAGCATGATCTGGGATTTGCTGGCTA 2100  
2101 CTATTATTCAGAGCGCCGACCAAAATATTTAAACAGGAATCTTAAGTAGGCGGT 2160  
2101 CTATTATTCAGAGCGCCGACCAAAATATTTAAACAGGAATCTTAAGTAGGCGGT 2160  
2161 ACAGTTTCCAAACACCTCTGAGAGGGATGATATCGCCCTCAATACACTTTCTGTC 2220  
2161 ACAGTTTCCAAACACCTCTGAGAGGGATGATATCGCCCTCAATACACTTTCTGTC 2220  
2221 CGGACAAATGAAAAATGATCCGAGAGGCCCATTTTGGGGGAAGTATCATTA 2280  
2221 CGGACAAATGAAAAATGATCCGAGAGGCCCATTTTGGGGGAAGTATCATTA 2280  
2281 TTACCTTACAAACAAGCTATCCAGATGAAATGTAACGGAATGGCGGATGCACTC 2340  
2281 TTACCTTACAAACAAGCTATCCAGATGAAATGTAACGGAATGGCGGATGCACTC 2340  
2341 CGAAGATTATTCGATTTGGAGGAGGAGGATTCGATCCGGAAGTATACCAATAAAA 2400  
2341 CGAAGATTATTCGATTTGGAGGAGGAGGATTCGATCCGGAAGTATACCAATAAAA 2400  
2401 TGGTGATC 2408  
2401 TGGTGATC 2408  
2401 TGGTGATC 2408

RESULT 2  
AR078182 2408 bp DNA Linear PAT 31-AUG-2000  
LOCUS AR078182  
DEFINITION Sequence 5 from patent US 5962297.  
ACCESSION AR078182  
VERSION AR078182.1 GI:10004928  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2408)  
AUTHORS Tsusaki,K., Kubota,M. and Chaen,H.  
TITLE Polypeptides having .beta.-fructofuranosidase activity  
JOURNAL Patent: US 5962297-A 5 05-OCT-1999;  
FEATURES  
1..2408  
Location/Qualifiers  
source /organism="unknown"  
BASE COUNT 777 a 492 c 496 g 643 t  
ORIGIN

Query Match 100.0%; Score 2408; DB 6; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 CGGGGAAAAATCTAGATTTCCAAATTTGGCCAGACTTCCAGTTGGTGTAGAGAGATTTCG 60  
1 CGGGGAAAAATCTAGATTTCCAAATTTGGCCAGACTTCCAGTTGGTGTAGAGAGATTTCG 60  
61 GACTCCCAATGACAGTGTGGCTAAGAAAACAGCTTACTCATGAGCAATTTACTAGAACAT 120  
61 GACTCCCAATGACAGTGTGGCTAAGAAAACAGCTTACTCATGAGCAATTTACTAGAACAT 120  
121 TTCAAAAGTCTGGGATTAAGGCCAAGTCCACTTTGAAATTAACCTTTAGCCTGCTGTG 180  
121 TTCAAAAGTCTGGGATTAAGGCCAAGTCCACTTTGAAATTAACCTTTAGCCTGCTGTG 180  
181 GGGGCTTTTGTGTTTATTTATTTCAACTGCAAGTGTCCATCCCTATATCAATTTAA 240  
181 GGGGCTTTTGTGTTTATTTATTTCAACTGCAAGTGTCCATCCCTATATCAATTTAA 240  
241 GACGAATTTCTAATCAATCCATGCCATCCCAATTAACCTGCTCTCTATCTTTAA 300  
241 GACGAATTTCTAATCAATCCATGCCATCCCAATTAACCTGCTCTCTATCTTTAA 300  
301 TTAATTAAGAACTATCAAGAGCTTTCTATCAATTAATTCATATTCGAAGGAGAGAG 360  
301 TTAATTAAGAACTATCAAGAGCTTTCTATCAATTAATTCATATTCGAAGGAGAGAG 360  
361 ATGAACCTTCAAAAGATTTGGCGAAAAAAGCAGCTGCCATCTTCAGAGCTGTATATTA 420  
361 ATGAACCTTCAAAAGATTTGGCGAAAAAAGCAGCTGCCATCTTCAGAGCTGTATATTA 420  
421 GTAGAGGAGGAGAGAGCCCATATTTTTCGCGCAATGAACAGCGGGGACTACAAAGAA 480  
421 GTAGAGGAGGAGAGAGCCCATATTTTTCGCGCAATGAACAGCGGGGACTACAAAGAA 480  
481 GACTATGTTTTGCCATATTTACAGCGCTGACATGCTAAATAATTCAGAGCAACAANAAC 540  
481 GACTATGTTTTGCCATATTTACAGCGCTGACATGCTAAATAATTCAGAGCAACAANAAC 540  
541 AGTCCTCAATTTAAAGTCCCTCAATTCATGATCAGCAATCAAAAACATGATTCGCA 600  
541 AGTCCTCAATTTAAAGTCCCTCAATTCATGATCAGCAATCAAAAACATGATTCGCA 600  
601 AAAGGATGATTAAGTCAAGTCAATTCATGATCAGCAATCAAAAACATGATTCGCA 660  
601 AAAGGATGATTAAGTCAAGTCAATTCATGATCAGCAATCAAAAACATGATTCGCA 660  
661 CAAAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
661 CAAAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
721 GGTGACCCCAAAAACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
721 GGTGACCCCAAAAACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
781 ACATCGATTGACAGTGGAAAAAATGCTGGAAGAGATTTGAAGATATGATGATGATGAT 840  
781 ACATCGATTGACAGTGGAAAAAATGCTGGAAGAGATTTGAAGATATGATGATGATGAT 840  
841 CCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
841 CCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
901 AAAGATGCGCAAGTCCGTTTATCTATACAGATTACTAGGTAATCTGAAGATGTTGGA 960  
901 AAAGATGCGCAAGTCCGTTTATCTATACAGATTACTAGGTAATCTGAAGATGTTGGA 960  
961 ACCGCTGCTGTATCAACCAATATTTCACTGCTCAAGTAACCTATATCCACCGGATGCA 1020  
961 ACCGCTGCTGTATCAACCAATATTTCACTGCTCAAGTAACCTATATCCACCGGATGCA 1020

```

OY 1021 GCTACACTTAAGATGATGAGTATCGATCATTAATCTGCTTTGATGCGGAGACGGT 1080
    |||||||
Db 1021 GCTACACTTAAGATGATGAGTATCGATCATTAATCTGCTTTGATGCGGAGACGGT 1080
OY 1081 ACAGTTTATCAAAATATTACAGCAATTTATCGATGAAGCGCAAGTGTTCAGTGATTAAC 1140
    |||||||
Db 1081 ACAGTTTATCAAAATATTACAGCAATTTATCGATGAAGCGCAAGTGTTCAGTGATTAAC 1140
OY 1141 CATACTTTAAGAGACCCCTCAGTATGTGAAGATGAAGCGCCATTAATATCTGTCTTTGAA 1200
    |||||||
Db 1141 CATACTTTAAGAGACCCCTCAGTATGTGAAGATGAAGCGCCATTAATATCTGTCTTTGAA 1200
OY 1201 GCGAATCTGGAACACAGATGTATGCAAGCGATCATCTTTCAATAATTAAGCTTAC 1260
    |||||||
Db 1201 GCGAATCTGGAACACAGATGTATGCAAGCGCGATCATCTTTCAATAATTAAGCTTAC 1260
OY 1261 TATGCGGAGAGTACGCTCTTCTCCAGATGAAGAAAAATAACTCTTCAAGTCTTAAA 1320
    |||||||
Db 1261 TATGCGGAGAGTACGCTCTTCTCCAGATGAAGAAAAATAACTCTTCAAGTCTTAAA 1320
OY 1321 AAACAAATTCCTTTAGCGAATGCTGATGATGATGATGATGATGATGATGATGATGAT 1380
    |||||||
Db 1321 AAACAAATTCCTTTAGCGAATGCTGATGATGATGATGATGATGATGATGATGATGAT 1380
OY 1381 ACAGTGAAGAGTGTATGAAACCATTAAGTGCATGAACACAGATGAGATGAATGCA 1440
    |||||||
Db 1381 ACAGTGAAGAGTGTATGAAACCATTAAGTGCATGAACACAGATGAGATGAATGCA 1440
OY 1441 CGCGCAATATATTTAAATGAATTAATGATGATGATGATGATGATGATGATGATGATG 1500
    |||||||
Db 1441 CGCGCAATATATTTAAATGAATTAATGATGATGATGATGATGATGATGATGATGATG 1500
OY 1501 AAAATGACAGATGATGATTAAGCAAAAGATGTTTATGCTGAGGCGCGGAGCGAC 1560
    |||||||
Db 1501 AAAATGACAGATGATGATTAAGCAAAAGATGTTTATGCTGAGGCGCGGAGCGAC 1560
OY 1561 TCCTTAATATGCGCACACACACCCGATTAATGAATGATGATGATGATGATGATGATG 1620
    |||||||
Db 1561 TCCTTAATATGCGCACACACACCCGATTAATGAATGATGATGATGATGATGATGATG 1620
OY 1621 GACCTGCTGATCTACACACACCTTACTCTCATGCGGATGATGCGGACCCCTGAAGTAT 1680
    |||||||
Db 1621 GACCTGCTGATCTACACACACCTTACTCTCATGCGGATGATGCGGACCCCTGAAGTAT 1680
OY 1681 AATGCTGCTACACAGTATATGACGAATGAGGCTTATCCGAACATCACTCTCAC 1740
    |||||||
Db 1681 AATGCTGCTACACAGTATATGACGAATGAGGCTTATCCGAACATCACTCTCAC 1740
OY 1741 CTGCGGAGACAGCTTGGGGTTAATATTAAGGGTCTGACACATCTGAGAGAGAAATAGT 1800
    |||||||
Db 1741 CTGCGGAGACAGCTTGGGGTTAATATTAAGGGTCTGACACATCTGAGAGAGAAATAGT 1800
OY 1801 TCCGGAACAGCAATTCCTCATAGGATTAATCTCCCAATTAAGAAAGTACAGCGCA 1860
    |||||||
Db 1801 TCCGGAACAGCAATTCCTCATAGGATTAATCTCCCAATTAAGAAAGTACAGCGCA 1860
OY 1861 GATCTACTCTTCCCAACCTTTTAAATAAAGGCTTTGTGACACCTTTAGAA 1920
    |||||||
Db 1861 GATCTACTCTTCCCAACCTTTTAAATAAAGGCTTTGTGACACCTTTAGAA 1920
OY 1921 GAAAAAAGATCTTGGCGCGGCAATTAACCATTTGCCGAAGAGACTACCAATTAANA 1980
    |||||||
Db 1921 GAAAAAAGATCTTGGCGCGGCAATTAACCATTTGCCGAAGAGACTACCAATTAANA 1980
OY 1981 AAAAAATTTTCCCTTTTAAAGGGGAGTCTTTTCTATCTTGGGGTGGGATTTGTTG 2040
    |||||||
Db 1981 AAAAAATTTTCCCTTTTAAAGGGGAGTCTTTTCTATCTTGGGGTGGGATTTGTTG 2040
OY 2041 GCGCCACAGGAGGACCTTTATTAATTTATGATGATGATGATGATGATGATGATGATG 2100
    |||||||
Db 2041 GCGCCACAGGAGGACCTTTATTAATTTATGATGATGATGATGATGATGATGATGATG 2100
OY 2101 CTATATTCAGGCGCGCGCAAAATATTAACCAAGAAATCTTAAGATGAGCGGT 2160

```

```

Db 2101 CTATATTCAGGCGCGCGCAAAATATTAACCAAGAAATCTTAAGATGAGCGGT 2160
OY 2161 ACAGTTTATCAAAATATTACAGCAATTTATCGATGAAGCGCAAGTGTTCAGTGATTAAC 2220
    |||||||
Db 2161 ACAGTTTATCAAAATATTACAGCAATTTATCGATGAAGCGCAAGTGTTCAGTGATTAAC 2220
OY 2221 CCGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
    |||||||
Db 2221 CCGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
OY 2281 TTACCTTTTAAACAAAGACTATTCAGATGAATGATGAATGATGATGATGATGATGATGATG 2340
    |||||||
Db 2281 TTACCTTTTAAACAAAGACTATTCAGATGAATGATGAATGATGATGATGATGATGATGATG 2340
OY 2341 CGAAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2400
    |||||||
Db 2341 CGAAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2400
OY 2401 TGGTGATC 2408
    |||||||
Db 2401 TGGTGATC 2408

RESULT 3
AR208488
LOCUS AR208488
DEFINITION Sequence 5 from patent US 6383769.
ACCESSION AR208488
VERSION AR208488.1 GI:21509655
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2408)
AUTHORS Tsusaki,K., Kubota,M., and Chaeen,H.
TITLE Polypeptides having .beta.-fructofuranosidase activity
JOURNAL Patent: US 6383769-A 5 07-MAY-2002;
FEATURES
    source 1..2408
    /organism="Unknown"
BASE COUNT 777 a 492 c 496 g 643 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 2408; DB 6; Length 2408;
Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 CGGGGAAAAATCTAGATTCCAAATTTGGCCAGACTTCCAGTTGGTGTGAAGAGAGTTGG 60
    |||||||
Db 1 CGGGGAAAAATCTAGATTCCAAATTTGGCCAGACTTCCAGTTGGTGTGAAGAGAGTTGG 60
OY 61 GACTGCCAATGACGCTGTGGCGTAAGAAAAACAGCTTACATGACGAATTTACTAGAAGAT 120
    |||||||
Db 61 GACTGCCAATGACGCTGTGGCGTAAGAAAAACAGCTTACATGACGAATTTACTAGAAGAT 120
OY 121 TTCAAAGTCCGCGGATTAAGGCCAAGTCCACTTTGAATTAACCTTTTACGCTCTGTG 180
    |||||||
Db 121 TTCAAAGTCCGCGGATTAAGGCCAAGTCCACTTTGAATTAACCTTTTACGCTCTGTG 180
OY 181 GGGGCTTTTGTGTTTATTTATTTCAACTGCAAGTGTGTCATCCCTATATCAATTTAA 240
    |||||||
Db 181 GGGGCTTTTGTGTTTATTTATTTCAACTGCAAGTGTGTCATCCCTATATCAATTTAA 240
OY 241 GACGAAATTTCAATCAATCCATGCGATCCCAATTAACCTGCTCTCTATATCTTTAA 300
    |||||||
Db 241 GACGAAATTTCAATCAATCCATGCGATCCCAATTAACCTGCTCTCTATATCTTTAA 300
OY 301 TTAATTAAGAACTATCAAGAGCTTTTATATCAAAATTTATATATATCAAGAGAGAGAG 360
    |||||||
Db 301 TTAATTAAGAACTATCAAGAGCTTTTATATCAAAATTTATATATATCAAGAGAGAGAG 360
OY 361 ATGAATCTTAAGAAATGATGGGCAAAAGACAGCTGCGGTACCTTCAGAGATGATATTA 420

```

|||||  
Db 361 ATAACTCTCAAAAGATTGGGAAAAAGAGCGCCGTAACCTTCAGGACTGCTAATTA 420  
QY 421 GTAGAGCGGACGCGACGATATTTTGGCAGCAATGAAACGCGGACCTCAAGGA 480  
Db 421 GTAGAGCGGACGCGACGATATTTTGGCAGCAATGAAACGCGGACCTCAAGGA 480  
QY 481 GACTATGTTTTGCCATATTTACAGCGCTGACATGCTAAATAATTCAGCAACAAC 540  
Db 481 GACTATGTTTTGCCATATTTACAGCGCTGACATGCTAAATAATTCAGCAACAAC 540  
QY 541 AGTCCCAATTTAAAGTGGCTCAATTCATGCATGCAATCAAAAACATTTGCGCA 600  
Db 541 AGTCCCAATTTAAAGTGGCTCAATTCATGCATGCAATCAAAAACATTTGCGCA 600  
QY 601 AAAGGATGATAGTACGACGACTTAATGATTTGATGGATGATGCTGACGCACTG 660  
Db 601 AAAGGATGATAGTACGACGACTTAATGATTTGATGGATGATGCTGACGCACTG 660  
QY 661 CAAAACGCTGATGCTGCTGCGCAAAATATCATGATATCATCATGCTGCTGCTAGCA 720  
Db 661 CAAAACGCTGATGCTGCTGCGCAAAATATCATGATATCATCATGCTGCTGCTAGCA 720  
QY 721 GGTGACCCAAAAACAGTGTATGATCTCCATTCATTTATTTATCAAAAAGTGTAT 780  
Db 721 GGTGACCCAAAAACAGTGTATGATCTCCATTCATTTATTTATCAAAAAGTGTAT 780  
QY 781 ACATGATTTGACAGCTGGAATAATGCTGGAAGATTTTGAATATGATTAATTTGT 840  
Db 781 ACATGATTTGACAGCTGGAATAATGCTGGAAGATTTTGAATATGATTAATTTGT 840  
QY 841 CCAAAATGATCCGATTTCTTAATATCAACACAGAGTGTGCTGCTGCTGCTTAAAC 900  
Db 841 CCAAAATGATCCGATTTCTTAATATCAACACAGAGTGTGCTGCTGCTGCTTAAAC 900  
QY 901 AAAGATGGCCAACTCCGTTATTTCTATACGATTACTCAGTAATCCTGAAAGTGTGA 960  
Db 901 AAAGATGGCCAACTCCGTTATTTCTATACGATTACTCAGTAATCCTGAAAGTGTGA 960  
QY 961 ACCGCTGCTGTAAACCAATCATTTCACTGCTCAAGTAACCTTATCCAGCGGATGA 1020  
Db 961 ACCGCTGCTGTAAACCAATCATTTCACTGCTCAAGTAACCTTATCCAGCGGATGA 1020  
QY 1021 GCTACACTTAAAGTCAGTATCTGATCATTAATCTGCTTGTGATGGCGAGAGCT 1080  
Db 1021 GCTACACTTAAAGTCAGTATCTGATCATTAATCTGCTTGTGATGGCGAGAGCT 1080  
QY 1081 ACAGTTTATCAAAATATTCAGCAATTTATGATGAAGGCAAGTGTATAGTGATAC 1140  
Db 1081 ACAGTTTATCAAAATATTCAGCAATTTATGATGAAGGCAAGTGTATAGTGATAC 1140  
QY 1141 CATACTTTAAGAGACCTCTACTATGTTGAAGATGAAGGCAATATCTGCTTTGAA 1200  
Db 1141 CATACTTTAAGAGACCTCTACTATGTTGAAGATGAAGGCAATATCTGCTTTGAA 1200  
QY 1201 GCGAATACTGGAACAACAGATGTTATCAAGGCGATCAGTCTTCAATAATAAAGCTTAC 1260  
Db 1201 GCGAATACTGGAACAACAGATGTTATCAAGGCGATCAGTCTTCAATAATAAAGCTTAC 1260  
QY 1261 TATGGCGGAAGTACGCTCTTCTTCAGAAATGAAAAAATAACTGCTTCAAGTCTTAA 1320  
Db 1261 TATGGCGGAAGTACGCTCTTCTTCAGAAATGAAAAAATAACTGCTTCAAGTCTTAA 1320  
QY 1321 AAACAATTTGCTCTTTAGGAATGCTATGAGCATTTGAGCATTTGCTTAAAGTCTAT 1380  
Db 1321 AAACAATTTGCTCTTTAGGAATGCTATGAGCATTTGAGCATTTGCTTAAAGTCTAT 1380  
QY 1381 ACAGTGAAGAAGTGTATGAACCATTAAGTCATCAACAACAGTAGCAGATGAAGTGA 1440  
Db 1381 ACAGTGAAGAAGTGTATGAACCATTAAGTCATCAACAACAGTAGCAGATGAAGTGA 1440  
QY 1441 CCGGCAATATATTTAAATGATATTAATGATCTATTCAGGATTCAGAGATTC 1500  
Db 1441 CCGGCAATATATTTAAATGATATTAATGATCTATTCAGGATTCAGAGATTC 1500

Db 1441 CCGGCAATATATTTAAATGATATTAATGATCTATTCAGGATTCAGAGATTC 1500  
QY 1501 AAATGACGATGATGAATTAACGACAAAGATGTTATATGCTAGAGCCCGGAGCGAC 1560  
Db 1501 AAATGACGATGATGAATTAACGACAAAGATGTTATATGCTAGAGCCCGGAGCGAC 1560  
QY 1561 TCCTTAAATGCGCCACACAACCCGATTAATGAATGAATGCACTTATTAATGAATCTT 1620  
Db 1561 TCCTTAAATGCGCCACACAACCCGATTAATGAATGAATGCACTTATTAATGAATCTT 1620  
QY 1621 GACCCCTGCTATCTCACACACACTTACTCTCATTTGGGGATTCGCCGACCTGAAGTAT 1680  
Db 1621 GACCCCTGCTATCTCACACACACTTACTCTCATTTGGGGATTCGCCGACCTGAAGTAT 1680  
QY 1681 AATGTGCTACTCAAGTATATGAGCAATAGAGGCTTATGACAGATCTGCTGAC 1740  
Db 1681 AATGTGCTACTCAAGTATATGAGCAATAGAGGCTTATGACAGATCTGCTGAC 1740  
QY 1741 CTGCGGACAAAGCTTGGGGTTAATATTAAGGCTTGACACATCTGAGAGAAATAGT 1800  
Db 1741 CTGCGGACAAAGCTTGGGGTTAATATTAAGGCTTGACACATCTGAGAGAAATAGT 1800  
QY 1801 TCCGAGACAGACAAATCCCATAGCGATATATCTCCCAATTAAGAAATGCTCACTGCA 1860  
Db 1801 TCCGAGACAGACAAATCCCATAGCGATATATCTCCCAATTAAGAAATGCTCACTGCA 1860  
QY 1861 GATCTCTACTTTTCCCAACCTTTTAAATAAAGGCTTTGACCACTTATGA 1920  
Db 1861 GATCTCTACTTTTCCCAACCTTTTAAATAAAGGCTTTGACCACTTATGA 1920  
QY 1921 GAAAAAAGAAATCTTGGCCCGGCAATTAACCAATGCGCAAGAGAGTACCAATAA 1980  
Db 1921 GAAAAAAGAAATCTTGGCCCGGCAATTAACCAATGCGCAAGAGAGTACCAATAA 1980  
QY 1981 AGAAATTTTCCCTTTTAAAGGGGAGTCTTTTCTATCTTGGGCTGGGATTTG 2040  
Db 1981 AGAAATTTTCCCTTTTAAAGGGGAGTCTTTTCTATCTTGGGCTGGGATTTG 2040  
QY 2041 GCCCCACAGGGGACCTTATTAATTAATGATGATGCGATTTGGCTGGGCT 2100  
Db 2041 GCCCCACAGGGGACCTTATTAATTAATGATGATGCGATTTGGCTGGGCT 2100  
QY 2101 CTTATTTATCCAAAGCCGCCAGCAAAATATTAACCAAGAAATCTTAAGATAGCGGT 2160  
Db 2101 CTTATTTATCCAAAGCCGCCAGCAAAATATTAACCAAGAAATCTTAAGATAGCGGT 2160  
QY 2161 ACAGTTTACCAACCTGCTGAGGGGATGATATGCGCTCAATACACTTTTCTGTT 2220  
Db 2161 ACAGTTTACCAACCTGCTGAGGGGATGATATGCGCTCAATACACTTTTCTGTT 2220  
QY 2221 CCGACAAATGGAATAATGATCCGACAGAGGCCATCTTTTGGGGAGATATCACTA 2280  
Db 2221 CCGACAAATGGAATAATGATCCGACAGAGGCCATCTTTTGGGGAGATATCACTA 2280  
QY 2281 TTACCTTTACCAAAAGACTATCCAGATGGAATGATGCGATGCGGATCAACGTC 2340  
Db 2281 TTACCTTTACCAAAAGACTATCCAGATGGAATGATGCGATGCGGATCAACGTC 2340  
QY 2341 CGAAGATTTATTCATTTGACGAGCAAGAGGATTTGCCATCCGAGATATACATAA 2400  
Db 2341 CGAAGATTTATTCATTTGACGAGCAAGAGGATTTGCCATCCGAGATATACATAA 2400  
QY 2401 TGGTGATC 2408  
Db 2401 TGGTGATC 2408

RESULT 4  
E15383 2408 bp DNA linear PAT 28-JUL-1999  
LOCUS Bacillus sp. gene for beta-fructofuranosidase, complete cds.  
DEFINITION  
ACCESSION E15383  
VERSION E15383.1 GI:5710066

KEYWORDS JP 199806586-A/2.  
SOURCE Bacillus sp.  
ORGANISM Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.  
REFERENCE 1 (bases 1 to 2408)  
AUTHORS Tsusaki, K., Kubota, M. and Chaen, H.  
TITLE POLYPEPTIDE HAVING BETA-FRUCTOFURANOSIDASE ACTIVITY  
JOURNAL Patent: JP 199806586-A 2 10-MAR-1998;  
HAYASHIBARA BIOCHEM LAB INC  
COMMENT OS Bacillus sp.  
PN JP 199806586-A/2  
PD 10-MAR-1998  
PE 09-JUN-1997 JP 1997164875  
PR 10-JUN-1996 JP 96P 170630  
PI TSUSAKI KEIJI, KUBOTA MICHIO, CHAEN HIROTO  
PC C12N15/09, C07H21/04, C07K14/32, C12N1/21, C12N9/10, C12N9/24, PC  
(C12N15/09, C12N1/07), (C12N1/21, C12R1:19), (C12N9/24, C12R1:19); CC  
Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..2408  
FT /organism='Bacillus sp.'  
FT /strain='V230'  
FT CDS 361..1824  
FT /product='beta-fructofuranosidase' FT  
sig\_peptide 361..456  
FEATURES  
source 1..2408  
location/Qualifiers  
BASE COUNT 777 a 492 c 496 g 643 t  
ORIGIN  
Query Match 100.0%; Score 2408; DB 6; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGGGAAATACAGATGATCCAAATGGCCAGACATCCAGTGGGTAGAGAAAGATCG 60  
DB 1 CGGGGAAATACAGATGATCCAAATGGCCAGACATCCAGTGGGTAGAGAAAGATCG 60  
QY 61 GACTGCCAATGCAGCTGTGGTGAAGAAACAGCTTACTCATGAGCAATTAAGAGAAAT 120  
DB 61 GACTGCCAATGCAGCTGTGGTGAAGAAACAGCTTACTCATGAGCAATTAAGAGAAAT 120  
QY 121 TTCAAAAGTCTGGGATGAAGCCCAAGTCCACTTTGAATTAACCTTTTCACCTCTGTGTG 180  
DB 121 TTCAAAAGTCTGGGATGAAGCCCAAGTCCACTTTGAATTAACCTTTTCACCTCTGTGTG 180  
QY 181 GGGGCTTTTGTATTTATTTATTTCAATGCAAGTGTCCATCCCTATATCAATTTAA 240  
DB 181 GGGGCTTTTGTATTTATTTATTTCAATGCAAGTGTCCATCCCTATATCAATTTAA 240  
QY 241 GACGAATTTCAATCAATCCATGCCATCCCAATTAACCTGCTCCTCTATACTTTAA 300  
DB 241 GACGAATTTCAATCAATCCATGCCATCCCAATTAACCTGCTCCTCTATACTTTAA 300  
QY 301 TTATATAGAACTATCAAGAGCTTTCTTATCAATTTATATATCAATCCAGAGAGAGAG 360  
DB 301 TTATATAGAACTATCAAGAGCTTTCTTATCAATTTATATATCAATCCAGAGAGAGAG 360  
QY 361 ATGAACTTCAAAAGATTTGGGAAAAAAGAGCTGGCTTAACCTTCAGACTGCTATTTTA 420  
DB 361 ATGAACTTCAAAAGATTTGGGAAAAAAGAGCTGGCTTAACCTTCAGACTGCTATTTTA 420  
QY 421 GTAGAGCGGACGACGCGATATTTTTCGACAGCAATGAACAGCGGAGCTACAAGAA 480  
DB 421 GTAGAGCGGACGACGCGATATTTTTCGACAGCAATGAACAGCGGAGCTACAAGAA 480  
QY 481 GACTATGCTTTTGGCCATATTAACAGCGGCTGACATGCTAAATTCAGAGACAACAAAC 540  
DB 481 GACTATGCTTTTGGCCATATTAACAGCGGCTGACATGCTAAATTCAGAGACAACAAAC 540

DB 481 GACTATGCTTTTGGCCATATTAACAGCGGCTGACATGCTAAATTCAGAGACAACAAAC 540  
QY 541 AGTCTTCATTTTAAGTGGCTCAATTCATGATTCAGCAATCAAAAACATGATTTGGCA 600  
DB 541 AGTCTTCATTTTAAGTGGCTCAATTCATGATTCAGCAATCAAAAACATGATTTGGCA 600  
QY 601 AAAGGATATATAGTACAGCAACTTAATGATTTAGATTAAGGATAGCTGGCCACTG 660  
DB 601 AAAGGATATATAGTACAGCAACTTAATGATTTAGATTAAGGATAGCTGGCCACTG 660  
QY 661 CAAAAGCTGATGCTGCTGCGCAATTTATCATGATATCATCTGCTCCCTTTAGCA 720  
DB 661 CAAAAGCTGATGCTGCTGCGCAATTTATCATGATATCATCTGCTCCCTTTAGCA 720  
QY 721 GGTGACCCAAAACAGTGTATGATCCCTCATTTATTTATCAAAAAGTCGGTAT 780  
DB 721 GGTGACCCAAAACAGTGTATGATCCCTCATTTATTTATCAAAAAGTCGGTAT 780  
QY 781 ACATGATGACAGCTGTGAAAAATGCTGAAGATATTTGAAGATATGATTAATTTGTT 840  
DB 781 ACATGATGACAGCTGTGAAAAATGCTGAAGATATTTGAAGATATGATTAATTTGTT 840  
QY 841 CCAATGATCGCTATCTTAAATATCAAAACAGAGAGTGTGCTGCTACTTTAAC 900  
DB 841 CCAATGATCGCTATCTTAAATATCAAAACAGAGAGTGTGCTGCTACTTTAAC 900  
QY 901 AAAGATGGCCAGCTGCTTATTTCTATACAGATTTACTCAGTAACTCCTGAAGTGTGGA 960  
DB 901 AAAGATGGCCAGCTGCTTATTTCTATACAGATTTACTCAGTAACTCCTGAAGTGTGGA 960  
QY 961 ACCGCTGCTGTAAACCAATCATTTCAACTGCTCAAGTAACCTTATCCAGCCGGATGA 1020  
DB 961 ACCGCTGCTGTAAACCAATCATTTCAACTGCTCAAGTAACCTTATCCAGCCGGATGA 1020  
QY 1021 GCTACACTTAAAGTCCATGAGATGATGATCATTAATCTGCTTTGATGGCGAGACGT 1080  
DB 1021 GCTACACTTAAAGTCCATGAGATGATGATCATTAATCTGCTTTGATGGCGAGACGT 1080  
QY 1081 ACAGTTTATCAAAATATTCAGCAATTTATGATGAAGGCAAGTGGATTCAGGTGATAC 1140  
DB 1081 ACAGTTTATCAAAATATTCAGCAATTTATGATGAAGGCAAGTGGATTCAGGTGATAC 1140  
QY 1141 CATACTTTAAGAGACCTCCTACTATGTTGAAGATAAAGGCCATTAATCTTCTTTGAA 1200  
DB 1141 CATACTTTAAGAGACCTCCTACTATGTTGAAGATAAAGGCCATTAATCTTCTTTGAA 1200  
QY 1201 GCGAATATCTGAACACAGATGTTATCAAGGGGATCAGTCTTCAATTAAGCTTAC 1260  
DB 1201 GCGAATATCTGAACACAGATGTTATCAAGGGGATCAGTCTTCAATTAAGCTTAC 1260  
QY 1261 TATGGCGAAGTGAAGTCTCTTCCAGAAATGAAGAAAAATTAACGCTTCAAGCTTAA 1320  
DB 1261 TATGGCGAAGTGAAGTCTCTTCCAGAAATGAAGAAAAATTAACGCTTCAAGCTTAA 1320  
QY 1321 AAACAAATTTCTTTAGCGAATGTTAGGCAATTTGATTTGCGCATAT 1380  
DB 1321 AAACAAATTTCTTTAGCGAATGTTAGGCAATTTGATTTGCGCATAT 1380  
QY 1381 ACAGTGAAGGTTATGAAGACCATTAAGTGCATCAAAACCATATACAGATGAAGTCGA 1440  
DB 1381 ACAGTGAAGGTTATGAAGACCATTAAGTGCATCAAAACCATATACAGATGAAGTCGA 1440  
QY 1441 CGGCGCAATATTTAAATGAATTAATGATATTTATTCAGGATTCAGAGATTC 1500  
DB 1441 CGGCGCAATATTTAAATGAATTAATGATATTTATTCAGGATTCAGAGATTC 1500  
QY 1501 AAATGACGAGTATGGAATTAAGCAAAAGATGTTATATGCTAGGCGCCGAGAGCAG 1560  
DB 1501 AAATGACGAGTATGGAATTAAGCAAAAGATGTTATATGCTAGGCGCCGAGAGCAG 1560  
QY 1561 TCCTTAATATGCGCCACACAAACCCGATTAATGAACAGTGTATTTGAACATGAATCTT 1620  
DB 1561 TCCTTAATATGCGCCACACAAACCCGATTAATGAACAGTGTATTTGAACATGAATCTT 1620

QY	1621	GACCCGCTGATCTACACACACTTACTCTCATTCGGGTATCCCGACCCCTGAAGGTAT	1680
Db	1621	GACCCGCTGATCTACACACACTTACTCTCATTCGGGTATCCCGACCCCTGAAGGTAT	1680
QY	1681	AATGTGTACTCACAAGTTATATGACGATAGAGGCTTCTATCCAGAACTACTCTCAC	1740
Db	1681	AATGTGTACTCACAAGTTATATGACGATAGAGGCTTCTATCCAGAACTACTCTCAC	1740
QY	1741	CTGGGGGACAGCTGGGGTAAATTTAAAGGCTGTGACATCTGGAGGAGAAATAGT	1800
Db	1741	CTGGGGGACAGCTGGGGTAAATTTAAAGGCTGTGACATCTGGAGGAGAAATAGT	1800
QY	1801	TCCGGACAGAGCAATTCCTACGATATATCTCCCAATAAAAGAAATGTCACGTGCAA	1860
Db	1801	TCCGGACAGAGCAATTCCTACGATATATCTCCCAATAAAAGAAATGTCACGTGCAA	1860
QY	1861	GATCCTTACCTTTTCCCAACCTTTTAAATAAAGGCTTTGTACACCTTTAGAA	1920
Db	1861	GATCCTTACCTTTTCCCAACCTTTTAAATAAAGGCTTTGTACACCTTTAGAA	1920
QY	1921	GAATAAAGAAATCTTGGCGGGGCAATTAACCAATGCGGAGAGCACTACCCATTAATA	1980
Db	1921	GAATAAAGAAATCTTGGCGGGGCAATTAACCAATGCGGAGAGCACTACCCATTAATA	1980
QY	1981	AGAAATTTTCCCTTTTAAAGGGGAGTCTTTTCTATCTTGGGGTGGGATTTGTTG	2040
Db	1981	AGAAATTTTCCCTTTTAAAGGGGAGTCTTTTCTATCTTGGGGTGGGATTTGTTG	2040
QY	2041	GCCCCCAGAGGAGCCTTTATTAATTAATGATGAAATCTGGCGATTTTCTGGGCTTA	2100
Db	2041	GCCCCCAGAGGAGCCTTTATTAATTAATGATGAAATCTGGCGATTTTCTGGGCTTA	2100
QY	2101	CTTATTTTCCAGGCGGCGCCCAAAATATTAACCAAGGAATCTGAGATGAGGGGT	2160
Db	2101	CTTATTTTCCAGGCGGCGCCCAAAATATTAACCAAGGAATCTGAGATGAGGGGT	2160
QY	2161	ACAGTTTACCAACACCTGCTGGAGGGATGATATCGCCCTCAATACCACTTTCTGTTTC	2220
Db	2161	ACAGTTTACCAACACCTGCTGGAGGGATGATATCGCCCTCAATACCACTTTCTGTTTC	2220
QY	2221	CCGACAAATGGAATAATGATCCGAGAGGCCCATCTTTTGGGGGAAGTATCATTAATA	2280
Db	2221	CCGACAAATGGAATAATGATCCGAGAGGCCCATCTTTTGGGGGAAGTATCATTAATA	2280
QY	2281	TTACCTTTACCAACAAAGACTATCCAGATGGAATGGTACGGAATGGGGCAATGCAAGCTC	2340
Db	2281	TTACCTTTACCAACAAAGACTATCCAGATGGAATGGTACGGAATGGGGCAATGCAAGCTC	2340
QY	2341	CGAAGATTTATTTGATGAGAGGAGGATTCCTCCGAAATATACCAATTAATAA	2400
Db	2341	CGAAGATTTATTTGATGAGAGGAGGATTCCTCCGAAATATACCAATTAATAA	2400
QY	2401	TGGTGATC 2408	
Db	2401	TGGTGATC 2408	

RESULT 5  
LOCUS AR078181 1365 bp DNA linear PAT 31-AUG-2000  
DEFINITION Sequence 4 from patent US 5962297.  
ACCESSION AR078181  
VERSION AR078181.1 GI:10004927  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1365)  
AUTHORS Tsusaki,K., Kubota,M. and Chaen,H.  
TITLE Polypeptides having .beta.-fructofuranosidase activity  
JOURNAL Patent: US 5962297-A 4 05-OCT-1999;  
FEATURES Location/Qualifiers

source	1. 1365	/organism="unknown"	
BASE COUNT	459 a	266 c 283 g 357 t	
ORIGIN			
Query Match	56.7%; Score 1365; DB 6; Length 1365;		
Best Local Similarity	100.0%; Pred. No. 4.6e-299;		
Matches 1365; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	457	ATGAACACGGGGGACTACAAAGAGACATGCTTTTGGCCATATTACACGGCGTGACATG	516
Db	1	ATGAACACGGGGGACTACAAAGAGACATGCTTTTGGCCATATTACACGGCGTGACATG	60
QY	517	CTAAAAATTCAGAGCAACAACACAGCTCCATATTAAAGGCTCAATTCATGATCA	576
Db	61	CTAAAAATTCAGAGCAACAACACAGCTCCATATTAAAGGCTCAATTCATGATCA	120
QY	577	GCAATCAAAAACATTGATTCGGCAAAAAGGATGTAAGTCAAGGCAACCTTAATGATTTA	636
Db	121	GCAATCAAAAACATTGATTCGGCAAAAAGGATGTAAGTCAAGGCAACCTTAATGATTTA	180
QY	637	GATGATGAGGATGATGCTGGCCACTGCAAAACCGCTGATGCTGCGGCAATTATCATGA	696
Db	181	GATGATGAGGATGATGCTGGCCACTGCAAAACCGCTGATGCTGCGGCAATTATCATGA	240
QY	697	TATCAGATGCTGCTCGCTTTAGCAGGTGACCCAAAAAACAGTGAATGATCTCCACTTCAT	756
Db	241	TATCAGATGCTGCTCGCTTTAGCAGGTGACCCAAAAAACAGTGAATGATCTCCACTTCAT	300
QY	757	TTATTTCTATCAAAAAGTGGTGAATATCATGATGATGACAGCTGGAATAATGCTGAAGATTA	816
Db	301	TTATTTCTATCAAAAAGTGGTGAATATCATGATGATGACAGCTGGAATAATGCTGAAGATTA	360
QY	817	TTTGAAGATATGATGATTTTGTTCCTCAATATGATCCGATCTTAAATATCAACACAGAG	876
Db	361	TTTGAAGATATGATGATTTTGTTCCTCAATATGATCCGATCTTAAATATCAACACAGAG	420
QY	877	TGTCAGGTTCTGCTACTTTTAACCAAGATGGCCAAAGCCGTTTATCTATACAGATTAC	936
Db	421	TGTCAGGTTCTGCTACTTTTAACCAAGATGGCCAAAGCCGTTTATCTATACAGATTAC	480
QY	937	TCAGTAACTCTGAGATGATGAGAACCGGTCTGTGTAACCAATCATTTCACTGCTCAA	996
Db	481	TCAGTAACTCTGAGATGATGAGAACCGGTCTGTGTAACCAATCATTTCACTGCTCAA	540
QY	997	GTAACCTATATCCAGCCGGATGCGACGCTACCTTAAAGTCGATGAGATTCGATCATTA	1056
Db	541	GTAACCTATATCCAGCCGGATGCGACGCTACCTTAAAGTCGATGAGATTCGATCATTA	600
QY	1057	TCTGCTTTGATGGCGAGAGCGTACAGTTATATCAAAATATTCAGCAATTTATGATGAA	1116
Db	601	TCTGCTTTGATGGCGAGAGCGTACAGTTATATCAAAATATTCAGCAATTTATGATGAA	660
QY	1117	GGCAAGTGATTTTCAAGTATACCATTAATTAAGAGACCCCTCACTATGTTGAAGATTAAG	1176
Db	661	GGCAAGTGATTTTCAAGTATACCATTAATTAAGAGACCCCTCACTATGTTGAAGATTAAG	720
QY	1177	GGCCATTAATATCTTGTCTTGAAGCGAATCTGGAACAAAGATGTTATCAAGGCGAT	1236
Db	721	GGCCATTAATATCTTGTCTTGAAGCGAATCTGGAACAAAGATGTTATCAAGGCGAT	780
QY	1237	CAGTCTTTCAATTAAGTCTTACATGCGGAGAGTACGCTCTCTCCAGAAAGAAAA	1296
Db	781	CAGTCTTTCAATTAAGTCTTACATGCGGAGAGTACGCTCTCTCCAGAAAGAAAA	840
QY	1297	AATAAAGCTTCAAAAGTCTTAATAAACAAATTCCTTTAGAGCATGATGAGGCT	1356
Db	841	AATAAAGCTTCAAAAGTCTTAATAAACAAATTCCTTTAGAGCATGATGAGGCT	900
QY	1357	ATTGTTGAATGGCGGATGATATACAGTGAAGGTTATGAACCAATTTATGTCATCA	1416
Db	901	ATTGTTGAATGGCGGATGATATACAGTGAAGGTTATGAACCAATTTATGTCATCA	960

QY	1417	AAACAGTAGAGATGAGTGAAGTGAACGGCCCAATATATTTAAATGAAATTAATTAATGAT	1478
DB	961	AAACAGTAGAGATGAGTGAAGTGAACGGCCCAATATATTTAAATGAAATTAATTAATGAT	1020
QY	1477	CTATTCACGAGATTGACAGAGATCCAAATGACAGAGTGAATTAACGACAAAGATGT	1538
DB	1021	CTATTCACGAGATTGACAGAGATCCAAATGACAGAGTGAATTAACGACAAAGATGT	1080
QY	1357	TATATGCTAGAGCCCGGAGGCGACTCTTAAATGGCCACACAAACCGATTAATGAAC	1598
DB	1081	TATATGCTAGAGCCCGGAGGCGACTCTTAAATGGCCACACAAACCGATTAATGAAC	1140
QY	1597	GGACTGTGATTTGAACATGATCTTGACCTGCTATGTCACACACACTGATCTCATATGC	1656
DB	1141	GGACTGTGATTTGAACATGATCTTGACCTGCTATGTCACACACACTGATCTCATATGC	1200
QY	1657	GGTATCCCGCACCCCTGGAAGTAAATGTTGGTACTACACAGTTATATGACCAATAGAGC	1718
DB	1201	GGTATCCCGCACCCCTGGAAGTAAATGTTGGTACTACACAGTTATATGACCAATAGAGC	1266
QY	1717	TTCTATCCAGAACATCATCTCTCACCTGCGGAGCAAGCTTGGGGTTAAATTAAGGTCT	1778
DB	1261	TTCTATCCAGAACATCATCTCTCACCTGCGGAGCAAGCTTGGGGTTAAATTAAGGTCT	1320
QY	1777	GACACATCTGGAGGAGAAATAGTTCCGGACAAAGACAAATTC	1821
DB	1321	GACACATCTGGAGGAGAAATAGTTCCGGACAAAGACAAATTC	1365
RESULT 6			
LOCUS	AR208487	1365 bp	DNA
DEFINITION	Sequence 4 from patent US 6383769.	Linear	PAT 20-JUN-2002
ACCESSION	AR208487		
VERSION	AR208487.1	GI:21509653	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1365)		
TITLE	Tsutsaki, K., Kubota, M. and Chaen, H.		
JOURNAL	Polypeptides having .beta.-fructofuranosidase activity		
FEATURES	Patent: US 6383769-A 4 07-May-2002;		
SOURCE	Location/Qualifiers		
	1..1365		
	/organism="unknown"		
BASE COUNT	459 a 266 c 283 g 357 t		
ORIGIN			
Query Match	56.7%;	Score 1365;	DB 6;
Best Local Similarity	100.0%;	Pred. No. 4.6e-29;	Length 1365;
Matches 1365;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps 0;
QY	457	ATGAACACCGGGGACTACACAGAGACATATGTTTGGCCATATATACACGGCTGACATG	516
DB	1	ATGAACACCGGGGACTACACAGAGACATATGTTTGGCCATATATACACGGCTGACATG	60
QY	517	CTAAATAATCCAGGACAAACAACAGTCTCAATTTAAAGTCCCTCAATTCATGATCA	576
DB	61	CTAAATAATCCAGGACAAACAACAGTCTCAATTTAAAGTCCCTCAATTCATGATCA	120
QY	577	GCAATCAAAAACATTTGATTCGGCAAAAGGATGATTAAGTACAGGCAACTTAATGATTTA	636
DB	121	GCAATCAAAAACATTTGATTCGGCAAAAGGATGATTAAGTACAGGCAACTTAATGATTTA	180
QY	637	GATATATGGATAGCTGGCCACTGCAAAACGCTGATGGTACTGGGCAATATATCATGA	696
DB	181	GATATATGGATAGCTGGCCACTGCAAAACGCTGATGGTACTGGGCAATATATCATGA	240
QY	697	TATACATCGTCTCCGCTTTAGCAGGTGACCCCAAAAAACAGTATGATACTCCACTCAT	756
DB	241	TATACATCGTCTCCGCTTTAGCAGGTGACCCCAAAAAACAGTATGATACTCCACTCAT	300

QY	757	TTATTCTATCAAAAAGTCGGTGATACATTCATTCGACAGCTGGAAAAATGCTGGAAAGCTA	816
Db	301	TTATTCTATCAAAAAGTCGGTGATACATTCGATTCAGCTGGAAAAATGCTGGAAAGCTA	360
QY	817	TTTGAGATATGATGATAAATTTGTTCCAAATGATCGGATCTTAATATATACAAACACAGAG	876
Db	361	TTTGAGATATGATGATAAATTTGTTCCAAATGATCGGATCTTAATATATCAACACAGAG	420
QY	877	TGCTCAGTCTGCTACTTTAACCAAGATGGCCAAGTCGCTTTATTCATACGATTAC	936
Db	421	TGCTCAGTCTGCTACTTTAACCAAGATGGCCAAGTCGCTTTATTCATACGATTAC	480
QY	937	TCAGTAATCCTGAAGATGCTGGAAACCGGTCCTGTAAACCAATCATTTCAACCTGCA	996
Db	481	TCAGTAATCCTGAAGATGCTGGAAACCGGTCCTGTAAACCAATCATTTCAACCTGCA	540
QY	997	GTAACCTTATCCACAGCCGATGACGCTACACTTAAAGTCGATGGAGATCTGATCATAA	1056
Db	541	GTAACCTTATCCACAGCCGATGACGCTACACTTAAAGTCGATGGAGATCTGATCATAA	600
QY	1057	TCTGCTTTGATGGCGGAGACGGTACGATTATCAAAATATTACGCAATTTATGATGAA	1116
Db	601	TCTGCTTTGATGGCGGAGCGGACGATCTTATCAAAATATTACGCAATTTATGATGAA	660
QY	1117	GGCAAGTGGATTTGACGATGATACCATACCTTTAAGACACCTCTCATGTGTTGAGATTAAG	1176
Db	661	GGCAAGTGGATTTGACGATGATACCATACCTTTAAGACACCTCTCATGTGTTGAGATTAAG	720
QY	1177	GGCCATAAATATCTTGTCTTTGAAGCGAATACGGAACACAGATGGTTATCAAGCGAT	1236
Db	721	GGCCATAAATATCTTGTCTTTGAAGCGAATACGGAACACAGATGGTTATCAAGCGAT	780
QY	1237	CAGCTCTTCATPATAAAGCTTACTATGCGGGAAGTGACGCTCTTCGCAATGAAAA	1296
Db	781	CAGCTCTTCATPATAAAGCTTACTATGCGGGAAGTGACGCTCTTCGCAATGAAAA	840
QY	1297	AATAAAGTGCCTTCAAGAGCTCAAAAAAACAATTCCTCTTACGGAATGGTGCTATTAGGC	1356
Db	841	AATAAAGTGCCTTCAAGAGCTCAAAAAAACAATTCCTCTTACGGAATGGTGCTATTAGGC	900
QY	1357	ATTGTGTAATGGCCGATGACTATACAGTGAAGAACTGTATGAACCACTTATGCGCATCA	1416
Db	901	ATTGTGTAATGGCCGATGACTATACAGTGAAGAACTGTATGAACCACTTATGCGCATCA	960
QY	1417	AACACAGTAGAGATGAAGTGAAGCGGCGCCATATATTTAAATGAATTAATATGCTAT	1476
Db	961	AACACAGTAGAGATGAAGTGAAGCGGCGCCCATATATTTAAATGAATTAATATGCTAT	1020
QY	1477	CTATTACGGATTCAGAGATCCAAAATGACGAGTGTGAATTAACGACAAAGATGT	1536
Db	1021	CTATTACGGATTCAGAGATCCAAAATGACGAGTGTGAATTAACGACAAAGATGT	1080
QY	1537	TATATGCTAGGAGCCCGGAGCGGACGCTCTTAAATGCGCCACACCAACCGGATTAATGAAGCT	1596
Db	1081	TATATGCTAGGAGCCCGGAGCGGACGCTCTTAAATGCGCCACACCAACCGGATTAATGAAGCT	1140
QY	1597	GGACTTGTATTAACATGATCTTGACCCGCTGATCTCACACACACTTACTCTCATTTGC	1656
Db	1141	GGACTTGTATTAACATGATCTTGACCCGCTGATCTCACACACACTTACTCTCATTTGC	1200
QY	1657	GGTATCCCCGACCCCTGAAGGTAAATATGTGGTACTCAACAAGTTATATGACGAATAGAGC	1716
Db	1201	GGTATCCCCGACCCCTGAAGGTAAATATGTGGTACTCAACAAGTTATATGACGAATAGAGC	1260
QY	1717	TTTATATCCAGACATCACTCTCACCTCGGGGACCAAGCTTGGGGTTTATATTAAGGCTCT	1776
Db	1261	TTTATATCCAGACATCACTCTCACCTCGGGGACCAAGCTTGGGGTTTATATTAAGGCTCT	1320
QY	1777	GACACATCTGGAGAGAAATATGTCCGGACAGACGAATTTCCCA	1821
Db	1321	GACACATCTGGAGAGAAATATGTCCGGACAGACGAATTTCCCA	1365

```
RESULT 7
ELI5382                               1365 bp   DNA       linear   PAT 28-JUL-1999
LOCUS      ELI5382
DEFINITION Bacillus sp. gene for beta-fructofuranosidase.
ACCESSION  ELI5382
VERSION     JP 199806586-A/1.
KEYWORDS   Bacillus sp.
SOURCE     Bacillus sp.
ORGANISM   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE
  1 (bases 1 to 1365)
  Tsusaki,K., Kubota,M. and Chaen,H.
  POLYPEPTIDE HAVING BETA-FRUCTOFURANOSIDASE ACTIVITY
  Patent: JP 199806586-A 1 10-MAR-1998;
  HAYASHIBARA BIOCHEM LAB INC

COMMENT
  OS   Bacillus sp.
  PN   JP 199806586-A/1
  PD   10-MAR-1998
  PE   09-JUN-1997 JP 1997164875
  PR   10-JUN-1996 JP 96P 170630
  PI   TSUSAKI KEIJI, KUBOTA MICHIO, CHAEN HIROTO
  PC   C12N15/09, C07H21/04, C07K14/32, C12N1/21, C12N9/10, C12N9/24, PC
  (C12N15/09, C12N1/07), (C12N1/21, C12R1:19), (C12N9/24, C12R1:19); CC
  strandedness: Double;
  CC   topology: linear;
  FH   key      Location/Qualifiers
  FT   source      1. 1365
  FT               /organism="Bacillus sp."
  FT               /strain="V230"
  FT   mat_peptide 1. 1365
  FT               /product="beta-fructofuranosidase"
  FT               Location/Qualifiers
  FT               1. 1365
  FT               /organism="Bacillus sp."
  FT               /db_xref="taxon:1409"

BASE COUNT  459 a 266 c 283 g 357 t

ORIGIN
Query Match      56.7%; Score 1365; DB 6; Length 1365;
Best Local Similarity 100.0%; Pred. No. 4.6e-299;
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ATGAACACGGGAGCTACAGAGAAGCTATGTTTGGCCATTTATACAGCGGCTGACATG 516
    |||||||
DB 1  ATGAACACGGGAGCTACAGAGAAGCTATGTTTGGCCATTTATACAGCGGCTGACATG 60
    |||||||

OY 517 CTAAAAATTCAGAGCAACAAACAGTCCCTCAATTTAAAGTGCCTCAATTTCATGCAATCA 576
    |||||||
DB 61  CTAAAAATTCAGAGCAACAAACAGTCCCTCAATTTAAAGTGCCTCAATTTCATGCAATCA 120
    |||||||

OY 577 GCAATCAAAAACATTGATTTGGGCAAAAGGATATGTAAGTCAAGCAACTTAATAGATTTA 636
    |||||||
DB 121  GCAATCAAAAACATTGATTTGGGCAAAAGGATATGTAAGTCAAGCAACTTAATAGATTTA 180
    |||||||

OY 637 GATGATGAGATAGCTGGCACTGCAAAAACGCTGATGAGTGGGCAAAATATGATGGA 696
    |||||||
DB 181  GATGATGAGATAGCTGGCACTGCAAAAACGCTGATGAGTGGGCAAAATATGATGGA 240
    |||||||

OY 697 TATCAATCTGTCCTCGCTTTAGCAGATGACCCCAAAAACAGTATGATGATCTCCACTTCA 756
    |||||||
DB 241  TATCAATCTGTCCTCGCTTTAGCAGATGACCCCAAAAACAGTATGATGATCTCCACTTCA 300
    |||||||

OY 757 TTATTTCTTCAAAAAGTGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 816
    |||||||
DB 301  TTATTTCTTCAAAAAGTGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
    |||||||

OY 817 TTTGAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876
    |||||||
DB 361  TTTGAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
    |||||||

OY 877 TGGTCAGGTTCTGCTACTTTTAAACCAAGATGCGCAAGTCCGTTTATTTCTATACAGATTAC 936
    |||||||
```

```
DB 421 TGGTCAGGTTCTGCTACTTTTAAACCAAGATGCGCAAGTCCGTTTATTTCTATACAGATTAC 480
    |||||||
OY 937 TCAGGTAATTCCTGAGAGATGCTGGAACCGGCTGCTGTAACCAAAATATTTCAACTGCTCAA 996
    |||||||
DB 481 TCAGGTAATTCCTGAGAGATGCTGGAACCGGCTGCTGTAACCAAAATATTTCAACTGCTCAA 540
    |||||||
OY 997 GTAACATATCCACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1056
    |||||||
DB 541 GTAACATATCCACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
    |||||||
OY 1057 TCTGCTCTTTGATGGGAGAGCGGTACAGTTTATCAAAATATTTCAAGATTTATTCAGATGAA 1116
    |||||||
DB 601 TCTGCTCTTTGATGGGAGAGCGGTACAGTTTATCAAAATATTTCAAGATTTATTCAGATGAA 660
    |||||||
OY 1117 GGCAGTGGATTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1176
    |||||||
DB 661 GGCAGTGGATTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
    |||||||
OY 1177 GGCATTAATATCTGCTCTTTGAGCGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1236
    |||||||
DB 721 GGCATTAATATCTGCTCTTTGAGCGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
    |||||||
OY 1237 CAGTCTTTCAATATTAAGCTTACTATGCGGAGTACGCTCTTCTCCAGATGAAAAA 1296
    |||||||
DB 781 CAGTCTTTCAATATTAAGCTTACTATGCGGAGTACGCTCTTCTCCAGATGAAAAA 840
    |||||||
OY 1297 AATAACGCTCTCAAAAGCTCTTAATAAACCAAAATGCTCTTTAGCGAATGCTGATTAAGC 1356
    |||||||
DB 841 AATAACGCTCTCAAAAGCTCTTAATAAACCAAAATGCTCTTTAGCGAATGCTGATTAAGC 900
    |||||||
OY 1357 ATTGTTGAATTTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1416
    |||||||
DB 901 ATTGTTGAATTTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
    |||||||
OY 1417 AACACAGTAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1476
    |||||||
DB 961 AACACAGTAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
    |||||||
OY 1477 CTATTCAGGATTTCAAGAGATGCTCAAAATGACGAGTGAATTAAGCACAAGATGTT 1536
    |||||||
DB 1021 CTATTCAGGATTTCAAGAGATGCTCAAAATGACGAGTGAATTAAGCACAAGATGTT 1080
    |||||||
OY 1537 TATATGCTAGAGCGCGGAGGAGCTCTTAATATGAGCCACACACACCGGTAATGAATGAACT 1596
    |||||||
DB 1081 TATATGCTAGAGCGCGGAGGAGCTCTTAATATGAGCCACACACACCGGTAATGAATGAACT 1140
    |||||||
OY 1597 GGACTTGTATTTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1656
    |||||||
DB 1141 GGACTTGTATTTGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
    |||||||
OY 1657 GGTATCCCGGACCGCTGAAGGATATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1716
    |||||||
DB 1201 GGTATCCCGGACCGCTGAAGGATATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260
    |||||||
OY 1717 TTTCTATTCAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1776
    |||||||
DB 1261 TTTCTATTCAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
    |||||||
OY 1777 GACACATCTGAGAGAAAAATAGTTCCGAGACAGCAATTTCCCA 1821
    |||||||
DB 1321 GACACATCTGAGAGAAAAATAGTTCCGAGACAGCAATTTCCCA 1365
    |||||||

RESULT 8
AE007686                               11318 bp   DNA       linear   BCT 27-JUL-2001
LOCUS      AE007686
DEFINITION Clostridium acetobutylicum ATCC824 section 174 of 356 of the
complete genome
ACCESSION  AE007686 AE001437
VERSION     AE007686.1 GI:15024739
KEYWORDS   Clostridium acetobutylicum.
SOURCE
```



ORGANISM Clostridium acetobutylicum  
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
Clostridium.  
1 (bases 1 to 11318)  
AUTHORS Nolling,J., Breton,G., Omeichenko,M.V., Markarova,K.S., Zeng,Q.,  
Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hiti,J., Wolf,Y.I.,  
Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,  
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.  
TITLE Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum  
J. Bacteriol. 183 (16), 4823-4838 (2001)  
MEDLINE 21359325  
JOURNAL 11466286  
PUBMED 2 (bases 1 to 11318)  
REFERENCE Childress,D., Zeng,Q. and Smith,D.R.  
AUTHORS Direct Submission  
TITLE Submitted (24-Jun-2001) GTC Sequencing Center Production,  
JOURNAL Flushing, and Bioinformatics teams, Genome Therapeutics Corp., 100  
Beaver Street, Waltham, MA 02453-8443, USA  
FEATURES  
source Location/Qualifiers  
1. 11318  
/organism="Clostridium acetobutylicum"  
/strain="ATCC 824"  
/db\_xref="ATCC:824"  
/db\_xref="taxon:1488"  
89. 790  
/gene="CAC1770"  
89. 790  
/codon\_start=1  
/product="RNA polymerase sigma factor, SigK-like"  
/protein\_id="AAK79735.1"  
/db\_xref="GI:15024740"  
/translation="MFMRSLEDRTOKAKESKEENRLOEKYPRIASLLQKTKCYL  
QYDDELSTGIAFEAEVESYEDNRKGFSLPAKVITLRAIDYRKQKRVNYLQ  
GNYSDNDINPIMFKAMEQYKNSINEIRLELEKKELEWGEIEFNVQASPK  
QEKLRKYKNIANYIVHODILNKLIRVKRLPIKEIKENIYINRKKLERGRYITAIIV  
IAEIGDYRLIREYE"  
801. 1544  
/gene="CAC1771"  
801. 1544  
/gene="CAC1771"  
/codon\_start=1  
/transl\_table=11  
/product="Uncharacterized protein, Ykrl B.subtilis  
homolog"  
/protein\_id="AAK79736.1"  
/db\_xref="GI:15024741"  
/translation="MKGIIEELKYLILTEKGFIEIDKLNTPRAKIGOEINIKSKK  
ISTKALREPALATAILFIIINSLIEMITREYVVDINPNKKNVAMEIHNYF  
GNITKLOANKSGIVKMKMKLKFATNVINNFIPRAONDKFIDSGKONTYITIA  
SSKSINDESIDSLSEHYIKENKINARPMIYIGNDYKSKQVPIPIDKTILINKVI  
KNNSPKYDILNKSIDELINIANDEMY"  
1682. 3151  
/gene="CAC1772"  
/note="sacB"  
1682. 3151  
/gene="CAC1772"  
/codon\_start=1  
/transl\_table=11  
/product="Levanucrase"  
/protein\_id="AAK79737.1"  
/db\_xref="GI:15024742"  
/translation="MNKLIYKCIILISGMISGIIITOOTFASNDNMKYPYVSHIT  
RYNSKIPMEQNDLAKFYPOPNASTLKIYASAKGIDKNGNLIDVDWSPFLQNDGT  
VANTGHITVIALGDPKNQDSTIYMYOKIGESIDSMNAGVFKVDSQYVANDP  
YLYKLOEMSGSATLISDQVRLFTYDFSGVAKDGSQVDSAQVITTTQVNSQPSNT  
INIDSVSHKSVFPGNGCTIYONVOQFIDEGKWSGDNHTLRDPHYVDNGRKLYLVE  
ANTGTNDYOGDSTLANKAFYGRSOSFEKTEKDLLIDTNKKHDSLANGLAIGTIELN  
NDYTLKREMKPLIASNTYDEIRANVFKKNGRWYLFPTDSGSKMTINGTISKOIYML  
GFSSNLTGPKYKPLNKTGLVNLNIDPDLFTYSHRAVPOTNGKNVITTSYITNRGM  
YSDHHSFAPSPFLNITKGTJKTJVSISNLLQOGQLTIDMY"

gene 3249. 4793  
/gene="CAC1773"  
/note="sacC"  
3249. 4793  
/gene="CAC1773"  
/codon\_start=1  
/transl\_table=11  
/product="Levanase"  
/protein\_id="AAK79738.1"  
/db\_xref="GI:15024743"  
/translation="MKRKKLKLITLITMFIYIILVLSLYIQNNSTIKSTTKKON  
REYVHFVYPNKKNNDPQRPVFEDEYHYLYLNQDIPNGNTEWRQATSKLVYWKOR  
GVCIIPKTYNENDIWSGSFVVAQNTAGFGALVATVQPSASMDQEOPLWYTKR  
GRTEKPYSNQPIINPCTKDPDRDKIIDEFKNNKVMVLAEGTKIGFESYNLKMOMH  
TGDEFTNIGIVEODPIEPMQSDNGNTKMIIGTSANGKVSGETPTVYAVWGYGKKE  
IADISPKMIDYGFDMVAAYVPESENGNGLVTPQYTSWLYVSGHDVLSFINEYFNG  
MDSIVREITLSEKQDNTYLSLSKPIKRENTITTSIDQFKQISVYGLKHLKQGYAYOL  
DTDINWDAKNVGLIRLESIDKRRHIDGIFTEKYSYVKNAYTGNPDKSKYVESRA  
PFDINNKVHLRIEVDKVSVEFIDCKITYSNEVFPREDKGTILFSIKGKAVFKNI  
TIKHID"  
4860. 6146  
/gene="CAC1774"  
/note="sacB2"  
4860. 6146  
/gene="CAC1774"  
/codon\_start=1  
/transl\_table=11  
/product="Levanucrase"  
/protein\_id="AAK79739.1"  
/db\_xref="GI:15024744"  
/translation="MKTRKTYKMTSLMVLAILTIFELLRHNTGYTSIWSROAOQ  
FKCIKENTAPNINPNFKLTAPNLWMDTPLVKKDGSIAVYNGKVFALIASRNVG  
NKRHDVAGISYECSDGEMNMYKGLAVVEDALGSROWAGSAIIDENGMQFETAG  
RKGEAVRREBORLYKTKRSINVDKGVAHITCSHOVLIEPDGYVQYOTMOAKPDIY  
SFROPYFREDKTKKDYLIIEGKNGGKIEKKPENIDKLERKHHIAPRGVENNGVY  
GIAVAONKDLTRFKLPLPDLAVGVNOOLEKPPQIVMKKNKYLETFISKFTYAGGLNG  
VDGLYGFQGNLSRNNYKPLNGNGCLVITNPIDPQYTSWLYVSGHDVLSFINEYFNG  
QLYRGYTFAPLQISLKGYSKIIIGRLEGVYTPAH"  
6560. 8305  
/gene="CAC1775"  
6560. 8305  
/gene="CAC1775"  
/codon\_start=1  
/transl\_table=11  
/product="Predicted membrane protein"  
/protein\_id="AAK79740.1"  
/db\_xref="GI:15024745"  
/translation="MKIKGIIAISMSTLLIGSISVYKASNNKRLONLYKEFQKL  
LNKDSNKSIVRTDNNKLSLDDLKNNKVIIPNFTILNKLKLPILIKSGDONSYSKTK  
GNSVSAAGDSKSNAPYAKLDPYILNBSLKNGETITNTQIAMIISDPTDAGDALSA  
VGSPSPAYILGVLDONNGEYVQFTNAGDFDLGYVDTSKGAKSNIVGSIHVIPEDY  
QAFGTQSTQSTYANVAVDESKMDAALIGTSGSKDYMNLKISDQSGNVQVWVYV  
DKKVVYINKPSSNAGICNYTTTITPTVLSNSTYNSRYLMIGKNNDYFAMLSGIONTI  
ILDKTYINENAKOPTOETPNKQMEFRTADGSTVITLLNNYPELRFQVROTEMLQVIL  
FTINDPNNASTHRSQYCGLANKYAEKRLNTTACQDYLVLYSPSLIQDPIINDETH  
LTVGDPIVLYDSTTAYADNYITGPTTGSSIANINVTNVPTAAVKDIYVASLNDVAR  
VSDFNMYKVAAPGESFMSDQOCNIGIDVYVYENGNNVQLMGTWRFVQAGVDPIS  
IIPGLHPYVEAGD"  
8693. 9346  
/gene="CAC1776"  
8693. 9346  
/gene="CAC1776"  
/codon\_start=1  
/transl\_table=11  
/product="HAD superfamily hydrolase"  
/protein\_id="AAK79741.1"  
/db\_xref="GI:15024746"  
/translation="MDRIILFDCYGTILDEKPVDAISIAKNNILNSKDAVNIYNY  
EDRLMVEGTIYPRYKLIKQVLEYQDMELVTNVEFSKSYDEITSVKRLKPEVEVESLE  
KIKKKGELIASVSHDIDYNDALGNVENNVEFSLSDIHAAYDOLKFFRYVEERLE  
LKQKHHIAGIYMWMDIYPCSKLQMKRIWVREKKKGNKCHERPEVHTLDEVINLI"  
9596. 10531  
/gene="CAC1777"



CDS 9598..10521  
/gene="CAC1777"  
/codon\_start=1  
/transl\_table=11  
/product="Nudix (MUT-1 like) hydrolase"  
/protein\_id="AAK79742.1"  
/db\_xref="GI:15024747"  
/translation="MRDMNKKNNKLEKEFELDRVYGVDERPSTVMDLFTVDVP  
IEKDPNKALEKILIKRGDHPYMGAVAGFVNIWEGISSAATYRELKEPTNENYF  
EOLKTEGDVNDPRPMRIVTSVAMALADKSTPKAGDDADKKTQVKKRFLSSGF  
GIDKIDITYNILLSDGETKIGLVTEKEKNGVVIKYPSFLLEMSKEELFDHE  
EDCALEKLNKLEIYPIAFSRLPKYFTLEAKQVEEAILNLEKPLTRANFRKIKM  
VTEKEKTSKIPATCYTFNEDMHTFLDE"  
10543..11193  
/gene="CAC1778"  
10543..11193  
/gene="CAC1778"  
/codon\_start=1  
/transl\_table=11  
/product="Amidase from nicotianamide family"  
/protein\_id="AAK79743.1"  
/db\_xref="GI:15024748"  
/translation="MKFPEKILNMESEIKGFVDLNDLKDXTAIVYDVMVGFHE  
GALSPRBEIVDEIVIKNEKILGNKRIFFLDEHTNNSFEKSYAHCEGSLAEELI  
PELKNALDLSNTVMIPKNSVNGFHAPEKMEENESQLENIIGCEVDIVSNFA  
NLKTYFNKNDKRIIIPSNVETDFGTHDGLMKIISLMEMOSNGIEIVDRVL"  
BASE COUNT 4466 a 1373 c 1945 g 3534 t  
ORIGIN  
Query Match 31.4%; Score 756.6; DB 1; Length 11318;  
Best Local Similarity 73.1%; Pred. No. 3.3e-161;  
Matches 985; Conservative 0; Mismatches 359; Indels 3; Gaps 1;  
QY 456 AATGAACGGGGGAGCTACAGAAGACATGTTTCCCATATTAACAGCCGTGCAT 515  
DB 1765 AACTAATGACATTAATTAAGAAACATATGAGATTTGCAATTAATCTGTTATATAT 1824  
QY 516 GCTAAAAATTCAGAGACACAAACAGTCCCTCAATTTAAAGTCCCTCAATTCAGATC 575  
DB 1825 GTCAAAAATTCAGATGAGCAAAATGACTTAAATTTAAAGTCCCTCAATTTAATGATC 1884  
QY 576 AGCAATCAAAACATGATTGCGCAAAAGGGATATGATAGTCAGCAACTTAATAGATT 635  
DB 1885 AACATTAATAAATATAGCTTACAGAAAGGGTATGATTAAGAAATGATTAATGATAGATT 1944  
QY 636 AGATGATGGAGATAGCGGCTGAGCAAAAGCGTATGATGCGGCAATATATCATG 695  
DB 1945 GGATGATGGAGATAGCTGGCCATTGCAAAATGGTATGAGCAAGTGGCTACTACATG 2004  
QY 696 ATATCAATCGTCCGCTTTAGCAGGTGACCCAAAACAGTATGATCTCCACTTCA 755  
DB 2005 GTATCAATATGTTTGGCTTTGGCAGGTGATCCTAAATAACAGATCATCTTATTTA 2064  
QY 756 TTATTTCTATCAAAAAGTCGCTATATCATGATGACAGCTGGAATAAGTCTGAGAGAT 815  
DB 2065 TATGTTTATCAAAAAGATTTGAGAAAACCTCATTGATAGTGAAGAATCAGGTAAGT 2124  
QY 816 ATTGAAGATGATGAATTTGTTCCAAATGATCCGTATCTAAATGAATGAACACAGA 875  
DB 2125 GTTTAAAGATGATGAATATACGTTGCCAATGATCCTTATCTTAAGATTAACACAGA 2184  
QY 876 GTGTCAGGTTCTGCTATTAACAAAGATGGCAAGTCTGTTATTTATATACAGATT 935  
DB 2185 ATGTCGTGTTCTGCAACTTAACATCTGATGTCAGAGTTCGTTTGTTCACAGATT 2244  
QY 936 CTAGGATATCTGAAGATGTTGAACCGGTGCTGTTAACAATCAATCTTCACTGCTCA 995  
DB 2245 TTGCGGAGTAGCAAAAGATGTTGAGACAGATGCTAGTAATCAAGTTATATACACTCTCA 2304  
QY 996 AGTAACCTTATCCAGCGGATGAGCTACACTTAAGTGAATGAGTATCTATCATCTAA 1055  
DB 2305 AGTAACCTTATCTCAACCAAGATTAATTAATGATGATGTAAGTATCATCTAA 2364

QY 1056 ATCTGCTTTGATGCGGAGACGCTAGCTGTTATCAAAATATTCAGCAATTTATCGATGA 1115  
DB 2365 ATCTGCTTTGATGCGGAGACGCTAGCTGTTATCAAAATATTCAGCAATTTATCGATGA 2424  
QY 1116 AGCAAGTGGATTTCAAGGTATTAACCTACTTTAAGACCCCTACTAGTTGAAGATTA 1175  
DB 2425 AGGTAATGAGATTTCTGGCATATATCATCTTGGAGATCTCATATATGTAAGATTA 2484  
QY 1176 GGGCCNAAATATCTGCTCTTTAAGAGATTAAGTGAACACAGATGTTATCAAGCGCA 1235  
DB 2485 TGGCCGCAAAATATCTGTTTGAAGCTTAACAGAGAAACAAAGTGTATCAACAGAGA 2544  
QY 1236 TCAGCTCTTCAATTAATTAACCTACTATGCGGAGAGTGGCTCTTCCAGAAATCAAA 1295  
DB 2545 TACTCTCTCTACTCAATTAACCTTTTATGTTAGAGTCAAGATTTCTTTAAACAGAAAA 2604  
QY 1296 AATAA--ACTGCTTCAAGTCCCTAATAAACAATAATGCTTCTTTAGCGAATGTCATT 1352  
DB 2605 AGATCAATTTATTAATGATCTAATAAAGCATGATGCTTCTTACCTAATGGGCGACT 2664  
QY 1353 AGGATGTTGAATGCGCGATGATCTATACAGTGAAGGATGTTATGAACCATTAAGTCG 1412  
DB 2665 AGGATTTATGAATTAATTAATGATGATTAATTAATTAATTAATTAATTAATTAATG 2724  
QY 1413 ATCAACACAGTAGCAGATGAAGTGAAGCGGCAATATTAATTAATGAATTAATGA 1472  
DB 2725 ATCAATACAGTAGCAGATGAAGTGAAGCGGCAATATTAATTAATGAATTAATGA 2784  
QY 1473 GTATCTATTCAGGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGAT 1532  
DB 2785 GTACTCTATTCAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGAT 2844  
QY 1533 TGTATATATCTAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1592  
DB 2845 TATATATATCTAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2904  
QY 1593 AACTGACATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1652  
DB 2905 AACTGACATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2964  
QY 1653 TTGCGGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1712  
DB 2965 TTTTGGCTGTTCTCAAACTAATGTTAATAAAGCTAGTAACTAATTAATTAATTA 3024  
QY 1713 AGGCTCTATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAG 1772  
DB 3025 AGGATGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAG 3084  
QY 1773 GTCTGACACATCTGAGAGAGAAATAG 1799  
DB 3085 AACCAAAACATCTGTTATTCACAAATAG 3111  
RESULT 9  
AX434229 1449 bp DNA linear PAT 28-JUN-2002  
LOCUS Sequence 2644 from Patent WO0229113.  
DEFINITION AX434229  
ACCESSION AX434229  
VERSION AX434229.1 GI:21659037  
KEYWORDS  
SOURCE Bacillus licheniformis.  
ORGANISM Bacillus licheniformis.  
REFERENCE 1. Berka, R. and Clausen, I. G.  
METHODS for monitoring multiple gene expression  
PATENT: WO 0229113-A 2644 11-Apr-2002;  
JOURNAL Novozymes Biotech, Inc. (US) : Novozymes A/S (DK)  
FEATURES  
source location/Qualifiers  
1..1449  
/organism="Bacillus licheniformis"  
/db\_xref="taxon:1402"  
BASE COUNT 492 a 341 c 313 g 303 t



```
/transl_table=11
/protein_id="CA26512.1"
/db_xref="GI:40119"
/db_xref="SPTREMBL:045663"
/translation="MCKGQSYRGTPIYF"
337..370
/note="Imp. Inverted repeat A'"
RBS
446..445
/note="Pot. rRNA binding site (sacB)"
CDS
464..1885
/codon_start=1
/transl_table=11
/product="Levansucrase"
/protein_id="CA26513.1"
/db_xref="GI:732568"
/db_xref="SWISS-PROT:P05655"
/translation="MNKKPKAFVLTFTALLAGCAQTOAFKETNPKYKYEYGIS
HITRDMLOIPEBOOKNEKYOVPEPDSSTIKNISKAGLDVDSPLQADSTVANTHG
YHIFALADPKNADDTSIYMEYQKVESTISDSWKNMGRIYKSDKEDANDSLIKDOT
QEMSGATFTSDGKIRLEFYDPSGKHYKQTLTAQVNVASDSLSLNTNGEDYKSIIF
DGDKTYQNVQOOFIDEQNSYSSGDHNTLRDPHYVEDKHKRYLVEFANTNGTEGYGSES
LEKAYYKSTSEFROSOXKLLQSDKKRTALANGALGMEILNDYTLKKYMKPLIAS
NTVTDLEIRANVEKMGKMYLFTDSRCKSMITDITSDIMLMYSVSLGPKPLIN
KTGLVLMKMDLDPNDVYTFYTSFAYPOAKGNVITSTMTNGFYADKQSTFAPSEFLN
IKGRISVVDSDILEGQQLTVNK"
slg_peptide
464..550
mat_peptide
551..1882
/misc_feature
1892..1928
/product="mature levansucrase (aa 1-444)"
repeat_unit
/note="region of dyad symmetry"
1892..1904
/note="inverted repeat B"
1916..1928
repeat_unit
/note="inverted repeat B'"
BASE COUNT
720 a 422 c 370 g 495 t
ORIGIN
Query Match 25.4%; Score 611.8; DB 1; Length 2007;
Best Local Similarity 65.1%; Pred. No. 2.9e-128;
Matches 1024; Conservative 0; Mismatches 492; Indels 57; Gaps 6;
```

```
QY 838 GTTCCAAATGATCCGTATCTTAAATATCAACACAGAGTGTGAGTTCTGCTACTTAA 897
D 1111111111111111111111111111111111111111111111111111111
Db 911 GATGCAAAATGATTTATCTTAAAGACCAACACAGAAATGTCAGGTTCAGCCACATTT 970
QY ACCAAAGATGGCCAGTCCGTTTATTCTATACAGATTACTAGTAATCTGAAGATGCT 957
D 1111111111111111111111111111111111111111111111111111111
Db 971 ACATCTGACGGAAAAATCCGTTTATTCTACACTGATTTCTCCGGTAAACATTA----- 1023
QY 958 GGAACCGGTGCGTGAACCAAAATCATTTCACTGCTCAAGTAATCTTCCACGCGGAT 1017
D 1111111111111111111111111111111111111111111111111111111
Db 1024 -----CGGAAACCAAACTGACAACTGCACAAAGTTAAAGTATACGATCAGAC 1072
QY 1018 GCAGCTACACTTAAAGTCGATGAGTATCTGATCAATTAATCTGCTTGGATGGCGAGAC 1077
D 1111111111111111111111111111111111111111111111111111111
Db 1073 AGCTCT---TTGAACATCAACGGGTAGAGATTATTAATCAATCTTTGA---CGGTAC 1126
QY 1078 GGTACAGTTTATCAAAATATTTCAGCAATTTATCCATGACGACAGCTGATTCAGTGAT 1137
D 1111111111111111111111111111111111111111111111111111111
Db 1127 GGAATAAGCTATCAAAATGTCACAGCAGTTTCATCGATGAAGCCACTACAGCTCAGCGCAC 1186
QY 1138 AACCTACTTTAAGAGACCCCTCAGTATGTTAGATTAAGGGCCATTAATATCTGCTTT 1197
D 1111111111111111111111111111111111111111111111111111111
Db 1187 AACCTACTGTCAGAGATCTCTACGTAGAGATTAAGGCCACAAATACTTATATTT 1246
QY 1198 GAAGCAATCTGTAACACAGATGTTATCAAGCGATCAGTCTTTCATTAATTAAGCT 1257
D 1111111111111111111111111111111111111111111111111111111
Db 1247 GAAGCAAACTGTAACCTGTAAGATGATGATCAAGCGCAAGAAATCTTTATTAACAAACA 1306
QY 1258 TACTATGCGGGAAGTGAAGCTCTTCTTCCAGAAATGAATAAATCTCTCAAGTCT 1317
D 1111111111111111111111111111111111111111111111111111111
Db 1307 TACTATGCAAAAGCAGCATCATCTTCTCCGTCGAAGAAATCAAAAATCTCTCAAGGCGAT 1366
QY 1318 AAAAACAATTTGCTCTTTAGCGAATGTCGATTAGCATTTGATTTGGCCGATAC 1377
D 1111111111111111111111111111111111111111111111111111111
Db 1367 AAAAAGCAGCGCTGAGTATGCAACAGCGCGCTCGGTATGATGACTTAACCATGAT 1426
QY 1378 TATACAGTGAAGAAGTGTATGAACCATTTAGTCGATCAACACAGTACGATGAATC 1437
D 1111111111111111111111111111111111111111111111111111111
Db 1427 TACACACTGAAGAAAGTATGAACCGCTGATGTCGATTAACACAGTAAAGTAATTT 1486
QY 1438 GACGCGCCCAATATATTAAATGAATTAATGTTCTTATTCAGGATTCAGAGGA 1497
D 1111111111111111111111111111111111111111111111111111111
Db 1487 GACGCGGGAACGCTTTAAATGAACGCAAAATGTCCTGTCACGACCTCCCGCGGA 1546
QY 1498 TCCAAATGACGAGATGATGAATTAAGCAAAAGTGTATTAATCTGAGGCGCGGAGGC 1557
D 1111111111111111111111111111111111111111111111111111111
Db 1547 TCAAAATGACGATTTAGCGCATTTACGTTAACGATATTTACATGCTTGTTATTTCT 1606
QY 1558 GACTCTTAAATGGCCACACAAACCCGATTAATGAAGTGAATGATTTGTAACATGAT 1617
D 1111111111111111111111111111111111111111111111111111111
Db 1607 AATTCTTTAAGTGGCCATTAACAGCGCTGAACAAACTGGGCTTGCTTAAATATGAT 1666
QY 1618 CTGACCTGCTGATCTACACACACTTACTCTATTCGCGATCCGACCCGATGAGGT 1677
D 1111111111111111111111111111111111111111111111111111111
Db 1667 CTGATCTCAAGATGATTAACCTTACTTACACACTTCTGCTGATCAAGCGAAAGA 1726
QY 1678 AATATGTTGTTACACAAAGTTATGACGAATAGAGCTTCTATCCAGACATCACTCT 1737
D 1111111111111111111111111111111111111111111111111111111
Db 1727 AACATATGCTGATTTACAGCTATATGACAAACAGAGATTTCTACGACAAACAAATCA 1786
QY 1738 CACTCGCGGACAAAGCTTGGGGTTAATTTAAAGGCTGTGACATCTGAGAGAAAT 1797
D 1111111111111111111111111111111111111111111111111111111
Db 1787 ACGTTTGGCCAGCTGCTCTCTGAAACATCAAGAGCAAGAAATCTGTGTCAAGAC 1846
QY 1798 AGTTCCGACAAAGCAATTCCTCATAGGATTTATCTCCCAATTAAGAAGATTCCTG 1857
D 1111111111111111111111111111111111111111111111111111111
Db 1847 AG---CATCTTGAACAAGCAATTAACAGTTAACAATTAAGAAAGCAAGAAATGC 1903
QY 1858 CAAGTCTTACTTTTCCCAACCTTTTAAATTAAGGGGTTTGTACCACTTTA 1917
D 1111111111111111111111111111111111111111111111111111111
Db 1904 CGATATCTTATTTGCAATTTCTTTATTTTATTCACATTAAGGTGATCCATATGAA 1963
QY 1918 GAAGAAAAAGAA 1930
```

Db 1964 CTATATAAAGCA 1976

RESULT 11  
LOCUS SYNPRLC 5941 bp DNA circular SYN 27-APR-1993  
DEFINITION Cloning vector (pRL278) for sacB-mediated positive selection for double recombinants in gram-negative bacteria.  
ACCESSION L05083  
VERSION L05083.1 GI:209132  
KEYWORDS Cloning vector DNA.  
SOURCE Cloning vector DNA.  
ORGANISM unidentified cloning vector  
REFERENCE 1 (bases 1 to 5941)  
AUTHORS Cai, Y.  
TITLE Molecular genetic approaches towards the understanding of heterocyst differentiation and pattern formation in the cyanobacterium *Anabaena* sp  
JOURNAL Unpublished (1992)  
REFERENCE 2 (sites)  
AUTHORS Black, T.A., Cai, Y. and Molk, C.P.  
TITLE Spatial expression and autoregulation of heter, a gene involved in the control of heterocyst development in *Anabaena*  
JOURNAL Unpublished (1992)  
FEATURES  
LOCATION/Qualifiers  
1..5941  
/organism="unidentified cloning vector"  
/db\_xref="taxon:45196"  
complement(473..1894)  
/note="Sach from B.subtilis: ORF"  
CDS  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA072307.1"  
/db\_xref="GI:209133"  
/translation="MNIKKRKAQATVLTFTTALLAGCAVQAPAKETNOKPYKEYYGIS  
HITRDMQLIPEQOKNEKYQVEFDSSTIKNISAKGLEWDSWPLQADGTVANVHG  
YHIFALAGDKNADDTISYMFYOKVETSIDSWKNAGRVKSDKFDANDSILKDOT  
QEMSGSATFTSDGRLILEYDSEGKHYKOTLTAOVNVSADSLNNGVEDKSI  
DGDCKTQONVOGFIDEGNSGSDNHTLBDPHYVDKKGKYLVEFANTGTEDGSGEES  
LPNKAYKGSQSPROESOKLLOSDKRTALALGALMILNDYTIKKYMKPLIAS  
NTVDEITERAVNFKNGKMYLFTDSRGSKMTIDGITSNDIYMLGVSNSLVGPYKPLN  
KTGLVLKMDLDNDVNTFYSHFAVPAQAGNNVITISYMTNGFYADKOSTPAPSELLN  
IKGRKTSYVKSILEOGSLIVNK"  
2749..3543  
/note="kanamycin and neomycin resistance determinant  
(nptII)"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA072308.1"  
/db\_xref="GI:209134"  
/translation="MIBDDGLHAGSPAWEVRLFGYDMAOQTIGSDAAYRSLNAGR  
PVLFTKDLSCALNELDEARLSMTATGPPCAVLDVTEAGRDWILLLEVGODL  
LSSHLPAREKYSIMADAKRLHTLDPATCPDHOAKHIERARTMEGLVDODLDE  
EHGGLAPAELEFARLKARMPDEEDLVTHGDCLPIMVENCRFSGFIDCGRLGVADRY  
QDILATRDIAEELGEGWADRFVLVYGAAPDSQRIAFVRLDEFF"  
3558..4815  
/note="multiple cloning sites"  
misc\_feature  
4815..5941  
/note="oriV and oriT genes from plasmid pMB1"  
BASE COUNT 1419 a 1329 c 1471 g 1722 t  
ORIGIN

Query Match 25.4% Score 611.8; DB 12; Length 5941;  
Best Local Similarity 65.1%; Pred. No.2.5e-128;  
Matches 1024; Conservative 0; Mismatches 492; Indels 57; Gaps 6;

QY 358 ACAGTGAATTCCTCAAAAGATGGCGAAAAACAGCTGCCGTAACCTTCAGAGCTGCTATA 417  
DB 1897 ACGATGACATCAAAAGATTTGCAAAACAGCAACAGATTAACCTTACGACGCACTG 1838  
QY 418 TTAGTAGAGGCGAGCGACCCATATTTTTCGCGCAGCAATGACAGCGGGGACTACAG 477

Db 1837 CTGGCAGCAGG---CGCAACTCAAGCGTTTCCGAAAGAAAGCAACCAACCCATATATAG 1781

QY 478 GAAGACATATGTTGGTTGGCCCATATATACAGCGCCGATGCTAAATAATTCGACGACAA 537  
Db 1780 GAACATAGCGCATTTTCCCATATTTACACCGCATATATGCTGCAAAATCCCGAACAGCA 1721

QY 538 AACAGCTCCATATTTAAAGTCCCTCAATCAATCAGCATCAGCAATCAAAACATGATTCG 597  
Db 1720 AAAATATGAANAATATCAAGTTCCGTAATTCATTCGTCACAAATTAATAATATCTCTT 1661

QY 598 GCAAAAGGCTATGATATAGTACAGCACTTAATAGATTAGATGATGGATAGCTGGCA 657  
Db 1660 GCAAAAGG-----CCGGAGGTTGGGAGACGCTGGCA 1628

QY 658 CTGCAAAAGCGTATGCTAGTGGGCAAAATATGATGATGATTCATCGTCCGCTTTA 717  
Db 1627 TTACAAAGCGCTGAGCGGCTGTGCAACATATACAGGCTACACATCGTCTTGGCATTA 1568

QY 718 GCAGGTGACCCCAAAACAGTATGATATCTCACCTCATTTATTTATCAAAAGTCGCT 777  
Db 1567 GCCGAGATCTTAATAAATGCGGATGACCATCGATTTACATGTTCTATCAAAAGTCGCG 1508

QY 778 GATACATGATTTGACAGCTGGAATAATGCTGGAAGATTTTGAAGATATGATTAATTT 837  
Db 1507 GAACTTCTATTTGACAGCTGGAATAACGCTGCGCGCTTTAAAGACAGCGCAAAATTC 1448

QY 838 GTTCCAAATGATTCGCTATCTAAATATCAAAACACAGGAGTGTGCTGCTACTTTA 897  
Db 1447 GATGCAATGATTTCTATCTTAAGAACCAACACAGATGCTAGGTTACGCCACATTT 1388

QY 898 ACCAAAGATGCGCAAGTCCGTTTATCTATACAGATTAATCTAGGTAATCTGAAGATGCT 957  
Db 1387 ACATCTGACGGAATAATCCGTTTATTTATCTACAGTGTCTTCGCGGAACATTA----- 1335

QY 958 GGAACCGTGTCTGTACCAATATCTTTCAACTGCTCAAGTAACTTATCCACCGCAT 1017  
Db 1334 -----CGCAAAACAAACCTGCAACTGACACAGTAAACGATACGATCAGAC 1286

QY 1018 GCAGTACACTTAAAGATGATGAGTATCATCATCAATATGCTTTGATGGGCGGAGC 1077  
Db 1285 AGCTCT---TTGAACATCAACGCTGTAGAGGATTAATCAATCTTTGA---CGGTGAC 1232

QY 1078 GGTACAGTTTATCAAAATATTCAGCAATTTATCATGATGAGGCAAGTGGATTCAGTGTAT 1137  
Db 1231 GGAATAACGTATCAAAATATGATCAGAGATTCATCATGATGAGGCAACTACAGCTCAGGCG 1172

QY 1138 AACCATCTTTAAGAGACCTCTACATATGTTGCAAGTAAAGCGCATTAATATCTGCTTT 1197  
Db 1171 AACCATGCTGAGAGATCTCTACATGATGAGTAAAGCCCAAAATACTTATGATATTT 1112

QY 1198 GAACGATACTGTGACAAACAGATGTTATGAAGCGGCTACGCTTTCATCAATAAATAGCT 1257  
Db 1111 GAGCAAACTGTGAGACTGGAAGATGCTACCAAGCGGAGAGATCTTTATTTAAACAAAGCA 1052

QY 1258 TACTATGCGGGAAGTGAAGTCTCTCTCCAGATGAATAAATAAATACTGCTTCAAGTCT 1317  
Db 1051 TACTATGCGCAAAACACATCATCTCTCCGTCAGAAAGTCAAAACCTCTCCAAAGCAT 992

QY 1318 AAAAAACAATTCCTTTTACCAATATGCTCATTAGGCAATTTGGAATTTGGCGGATGAC 1377  
Db 991 AAAAAACCGCAGCTGATGATGACAAAGCGGCTCTCGGTATGATGATGATGATGATGAT 932

QY 1378 TATACAGTGAAGAAGCTTATGAACCATTAATGATGATGATGATGATGATGATGATGATG 1437  
Db 931 TACACACTGAAAAAAGATATAAACCGCTGATGATGATGATGATGATGATGATGATGATG 872

QY 1438 GAACGCGCAATATATTTAAATGAATTAATGATGATGATGATGATGATGATGATGATGATG 1497  
Db 871 GAACGCGGAGACGCTTTAAATGAACGCGCAATGATGATGATGATGATGATGATGATGATG 812

QY 1498 TCCAAAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1557

```

Db      811 TCAAAATGAGCATGAGCGATTACGTCTAAGCATATTACATCTGTATGTCT 752
QY      1558 GACTCTTAATATGCGCCACACACACCCGATTAATGAACTGCTGTATGACATGAT 1617
Db      751 AATCTTTTAAGTGGCCATACAGCGCTGAACAAACTGGCTGTGTAAATGAT 692
QY      1618 CTTCACCTGCTGATCATACACACACTACTCTCATCTGATGATCCGACACCTGAAGT 1677
Db      691 CTGTACTCTAAGCATGATACCTTTTCTACTACACTCTGCTGATCTCAGCGAAGGA 632
QY      1678 AATAATGTGTACTCACAAGTTATATGACGATAGAGGCTTCTATCAGAACATCATCT 1737
Db      631 AACATGTCTGATATGACGATATGACAAACAGAGATCTCAGACAAACATCA 572
QY      1738 CACCTGGCGGACAGCTGGGTTATATATTAAGGCTGACACATCTGAGAGAAAT 1797
Db      571 AGTTTGGCGGACAGCTTCTCTGACATCAAGGAAAGAAACATCTGTCTCAAAAC 512
QY      1798 AGTTCCGAGACAGCAATTCCTCCATGAGGATATCTCCCAATAAAGAAATGCTACTG 1857
Db      511 AG---CATCTTGAACAGACAAATTACAGTTAACAAATAAACGCAAAAGAAATGC 455
QY      1858 CAAGATCTTACCTTTTCCCAACCTTTTATTAATTAAGGTTTGTGACCACTTTA 1917
Db      454 CGATATCTTATGCGATTTCTTTATTTCTTATCATACATAAAGGTGAATCCCATATGA 395
QY      1918 GAAGAAAGAAAGAA 1930
Db      394 CTATATTAAGCA 382

RESULT 12
SYNPR/A/C
LOCUS   SYNPR/A 6347 bp DNA circular SYN 27-APR-1993
DEFINITION
Cloning vector (pRL271) for sacB-mediated positive selection for
double recombinants in gram-negative bacteria.
ACCESSION
105081
VERSION
105081.1 GI:209125
KEYWORDS
Cloning vector.
SOURCE
Cloning vector DNA.
ORGANISM
"unidentified cloning vector
artificial sequences: vectors.
1 (bases 1 to 6347)
Cal, Y.
REFERENCE
1 Molecular genetic approaches towards the understanding of
heterocyst differentiation and pattern formation in the
cynobacterium Anabaena sp
unpublished (1992)
2 (sites)
Black, T.A., Cal, Y. and Wolk, C.P.
Spatial expression and autoregulation of hetR, a gene involved in
the control of heterocyst development in Anabaena
unpublished (1992)
JOURNAL
FEATURES
source
Location/Qualifiers
1..6347
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
complement(473..1894)
/note="sacB gene from B. subtilis; ORF"
/codon_start=1
/transl_table=11
/protein_id="AA72302.1"
/db_xref="GI:209126"
/translation="MNIRKFAKQATVLTFTTALLAGTQAFKETNPKPKETIYGIS
HITRDMQLQIEQOKNEKYQVPEPDSSTIKINISAKGLVWDWELQADGIVANYHG
HYIFALAGDKNADDTISIMFYOKVETSIDSMNAGRVRSDKFPDANSILKDOT
QWMSGATFTSDGKIRLETFDSFKHYKQULTTQAVNVASDSSLNNGEDYSIF
DGDGTQYNOVOFFIDEGNYSGDNHTLRDPHYEDGKHYVEFANTGTGEOYGEES
LFNKAYGKSTSPROESOKILQSDPKRTTALANGALIMELNDYTKKWKPLIAS
NTVDEIRAVNPKNGKRWLYLFTDSRSGKMTIDGTTSDITRLGLVSNLGPRLPLN
KTGLVLRKDLDPNDVTFYSHFAVPAQAGNNVITSYMTNRGFTYADKOSTAPSPFLN
IKGKTSYVXKDSILEQGLTVNK"
complement(2393..3127)
CDS

```

```

/note="erythromycin resistance determinant"
/codon_start=1
/transl_table=11
/protein_id="AA72303.1"
/db_xref="GI:209127"
/translation="MNEKNIKHSQNFITSKHNIDKIMTINIRLNEHDNIFEIGSGKHF
TLELVKRCNFPVTAIEIDHKLCKTETENKLVVDNPNVLRKDLLOEPFKNOSQYKIGNI
PYNISTDIIRKIVPDSINLELYLVEGFAVDNPNVLRKDLLOEPFKNOSQYKIGNI
EYFHRKPRVNSLSILSRKSRISIKDKOKNTYFPMKVNKEKKIFTKNOFPNNSLKH
AGIDDLNNISEQFLSNFYKLEPK"
3503..4162
/note="chloramphenicol resistance determinant"
/codon_start=1
/transl_table=11
/protein_id="AA72304.1"
/db_xref="GI:209128"
/translation="MEKRTGTYTVTDISQMRKHEFAFQSAOVCYNOTVOLDITPAF
LTVKRNKRKHEFPAPFIHLARLNANHPFRAMKMGDELIVDSIPCYVHEDETER
LSLMEYHDFDROFLHIYSODVACYGENLAVEPKGIENMFVSANPMVSTSPDLNV
ANMDFFAPVFTMGKTYTGSDKVLMPALQYHHAACDFPHYGRMLNELQYCDENQGS
A"
4252..5221
/misc_feature
/note="multiple cloning sites"
misc_feature
5222..6347
/note="oriV and oriT genes from plasmid pMB1"
BASE COUNT 1649 a 1305 c 1404 g 1989 t
ORIGIN
Query Match 25.4%; Score 611.8; DB 12; Length 6347;
Best Local Similarity 65.1%; Pred. No. 2.5e-128;
Matches 1024; Conservative 0; Mismatches 492; Indels 57; Gaps 6;
QY 358 ACGATGAACCTGTAAGAATGGCGGAAAGACAGCTGCCGTAACCTTACAGCACTCTATA 417
Db 1897 ACGATGAACCTGTAAGAATGGCGGAAAGACAGCTGCCGTAACCTTACAGCACTG 1838
QY 418 TTAGTAGAGGCGGACGACCGCATATTTTGGCGAGCAATGACAGGGGAGCTACAG 477
Db 1837 CTGCGAGGAGG---CGCACTCAACCGTTTGGCAAGAAACGACAAAGCATATAG 1781
QY 478 GAAGACTATGCTTTTGGCCATATTTACAGCGCGTGAACATTAATAATTCACAGACAGCA 537
Db 1780 GAACATATGAGGATTTCCATATTTACAGCGCATATGATGATGATGATGATGATGATG 1721
QY 538 AACAGTCTCAATTTAAAGTGGCTCAATTCATGATGATGATGATGATGATGATGATG 597
Db 1720 AAAAATGAAGAAATATCAAGTTCCTGATTCGATTCGATTCGATTCGATTCGATTC 1661
QY 598 GCAAAAGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657
Db 1660 GCAAAAGG-----CTGAGAGCTTTGGGACAGCTGGGCA 1628
QY 658 CTGCAAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 717
Db 1627 TTACAAAGCGTGAAGGACAGCTGCGCAAACTATCAGCGGTACACATGCTTTGCATTA 1568
QY 718 GCAGGTGACCCAAAAAAGATGATGATGATGATGATGATGATGATGATGATGATG 777
Db 1567 GCCGAGATCTCAAAAGCGGATGATGATGATGATGATGATGATGATGATGATGATG 1508
QY 778 GATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 837
Db 1507 GAACTTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1448
QY 838 GTTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 897
Db 1447 GATGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1388
QY 898 ACCAAGATGCGGACAGTCCGTTATTCATGATGATGATGATGATGATGATGATGATG 957
Db 1387 ACATGTCGCGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1335
QY 958 GAACCGGCTGCTGATACCAATTCATTCATGATGATGATGATGATGATGATGATGAT 1017

```



Db	1720	AAAAATGAAATATATCAAGTTCCTGGAATTCGATTCGCCAATTTAAATATATCTTTCT	1661
OY	598	GCAAAAGGGTATGATTAAGTCAGGCACCTTAATPAGATTTAGATGATGGAGATAGCTGGCCA	657
Db	1660	GCAAAAGG-----CCGAGAGTTTGGAGACCTGGCGCA	1628
OY	658	CTGCAAAAACGGTGAATGGTACTGGCGCAATTATCATGATATATCAATTCGTCCGCTTTTA	717
Db	1627	TTTACAAAACGGTGAACGGCAGCTGTGGCAAACTATACAGCGCTACACATTCGTCCTTGGCATTA	1568
OY	718	CGAGGTGACCCAAAACAAACAGATGATPACTCCATCTTCAATTAATCTATCAAAAAGTGGGT	777
Db	1567	GCCGGAGATCTCTAAAAATGGGATGACACATCGATTTTCATGTTCTTATCAAAAAGTGGGC	1508
OY	778	GATACATCGATTGACAGCTGGAAAAATGCTGGAAAGATATTTGGAAGATATGATTAATTT	837
Db	1507	GAACCTTCTATTGACAGCTGGAAAAACGCTGGCGCGCTTTAAAGACAGGCACAAATTC	1448
OY	838	GTTCCAAATGATCCGCTATCTTAATATCAACACAGAGTGGTCAAGTCTTGCTACTTTTA	897
Db	1447	GATGCAAAATGATTTTATCTTAAAAAGACCAAAACAGAAATGGTCAAGGTCAGCCACATTT	1388
OY	898	ACCAAGATGGCCCAAGTCCGTTTATTCATACATCTACAGTACTAGGTAATCCGGAAGATGGT	957
Db	1387	ACATCTGACGAAAAAATCCGTTTATTCTACACTATTTCTCCGGTAAACATTA-----	1335
OY	958	GGAACCGGTGCTGGTAACCAATCATTTCAACTGCTCAAGTAAACTTATCCACCGCGAT	1017
Db	1334	-----CGGCAAAACAAACACTGCACACTGCACAGTTAAAGTATCAGCATCAGAC	1286
OY	1018	GCACCTACACTTAAAGTCGATGGAGTATCTGATCATAAATCTGCTTTGATGGCGGAGAC	1077
Db	1285	AGCTCT--TTGAACATCTAACGGGTGAGAGATTAATATCAATCTTTGA--CGGTGAC	1232
OY	1078	GGTACAGTTTATCAAAATATTCACCAATTTTTCATGAAGGCAAGTGGATTTCCAGTGGAT	1137
Db	1231	GGAAAAACGTTCAAAATGTTACACACAGTTCATCATGAAGCACTACAGCTCAGGCGAC	1172
OY	1138	AACCATCTTTAAGAGACCCCTCACTATGTTGAAGATAAGGCGCAATAATCTTGCTTT	1197
Db	1171	AACCATAGCGTGAGAGATCTCACTACGTAGAAAGATPAAAGGCCCAAAATACTTGATTT	1112
OY	1198	GAACCGAATACTGAAACACAGATGTTATCAGGCGATCAGTCTTCATATATPAAAGCT	1257
Db	1111	GAACCAAAACACTGGAACATGGAAGTGGGTACCAAGCAAGAACTTTATTAACAAAGCA	1052
OY	1258	TACTATGGCGGAAGTGAAGCTCTCTCCAGATGAAAAAATAAAGCTTCAAAAGCTT	1317
Db	1051	TACTATGGCAAAACACATCATCTTTCTCCGTCAAAAGTCAAAACCTTCCGAAGCGAT	992
OY	1318	AAAAAACAAATGCTCTTTTAGCAATAGTGCATTAGGATGTTGAATTTGGCGGATGAC	1377
Db	991	AAAAAAGCGACGGGTGAAGTGAACAAAGCGGCTCTCGGTATGATTTGAGCTAAACGATGAT	932
OY	1378	TATACAGTGAAGAAGTGTATGAACAATTAAGTGCATCAACACAGTAAAGCATGAAAGTC	1437
Db	931	TACACACTGAAAAAAGTATGAACACCGGTGATTCATCTAACACAGTAAAGATTAAT	872
OY	1438	GAAGCGCGCATATATTTAAATTAATTAATTAATGATATCATTCACAGGATTTCAAGAGA	1497
Db	871	GAAGCGCGGAACGCTTTTAAATTAACGGCAAAATGGTAAACCTGTTCACTCCGCGGCA	812
OY	1498	TCCAAAATGACAGATGATGAATTAAGCACAAAGATGTTTATATGTAAGGCGCGGAGGC	1557
Db	811	TCAAAATGACGATTTAGCGCATTAACGTTACGATATTTTACATGCTTTGTTATGTTTCT	752
OY	1558	GACTCCTTAAATGGCCACACACCCGATTAATGAACCTGACCTGTAATGCAACGAT	1617
Db	751	AATCTTTAATCGCCCATACAAACCGCGTGAACAAACTGGCCTTGTTTAAATGGAT	692
OY	1618	CTTGAACCTGCTGATTCACACACACTTACTCTATTCGCGGTATCCCGCACCTGGAAGGT	1677
Db	691	CTTGAATCTTAACGATAGTAACTTTACTTACTACACTTCGCTGATCCCAAGCAAGAGA	632

QY	1678	AATAATGGGATCCACAAAGTTATATATACCAATAGAGCTTCTATCCAGAACATCACTCT	1737
Db	631	AACATATGCTGGATTACAGCTATATATGACAAACAGAGATTTCTACGACGACAAACATFCA	572
QY	1738	CACCTGCGGGAAGAGCTTGGGGTTAAATATTAAGAGGCTGCACACATCTGGAGAGAAAT	1797
Db	571	ACGTTTGGGCCACAGCTTCTCTCTGACATCATCAAGAGCAAGAAACATCTGTTGTCAAGAC	512
QY	1798	AGTTCCGGACACAGCAATTCCTCCATAGCGATTATCTCCCAATTAAGAAATGTCACTGG	1857
Db	511	AG---CATCTCTGGAACAAGACCAATTAACAGTTTACAAATTAAGAAACCAAGAAATATGC	455
QY	1858	CAAGATCTCTACTTTTCCCAACTTTTTTTAAATTAAGGTTTGTGACCACTTTA	1917
Db	454	CGATTCCTTATGGCGATTTTCTTTATTTCTTATTCACATTAAGGTGAATTCATATGAA	395
QY	1918	GAAGAAAAAGAA	1930
Db	394	CTATATAAAGCA	382
RESULT 14			
LOCUS	AF048702/c	9297 bp	DNA
DEFINITION	Site-specific excision vector pFLP2, complete sequence.		
ACCESSION	AF048702		
VERSION	AF048702.1	GI:2961152	
KEYWORDS	site-specific excision vector pFLP2, site-specific excision vector pFLP2 artificial sequences: vectors.		
SOURCE	ORGANISM		
REFERENCE	AUTHORS		
TITLE			
JOURNAL			
FEATURES			
rep_origin			
gene			
CDS			
gene			
CDS			
gene			
CDS			

[illegible]

D	b	3762	ACGATGAAACATCAAAAACCTTTGCCAAAACACACAGATTTAACTTTACTACCGCACTG	3703
O	y	418	TTAGTAGAGCGGACGACGCGCATATTTTTCGCGACGAATGAACAGCGGAGCTACAG	477
D	b	3702	CTGCGCAGAGG --- CGCAACTCAAGCGTTTTGCGAAAGAACGAACCAAAACCCATTAA	3646
O	y	478	GAACAGTATGGTTTTGGCCATATTAACCGGCGCAACATGCTAAAAATTCCGAGCAACAA	537
D	b	3645	GAACCATATGCGCATTTCCCATATTACACGCGCATATATGCTGCANAATCCCGAACAGCA	3586
O	y	538	AACAGTCCTCAATTAAAGTGCCTCAATTCAATGCATCAGCAATCAAAAACATTGATTG	597
D	b	3585	AAAAATGAAAAATATATCAAGTTCCGTAGTTCGATTGTCACAAATTAATAAATATCTCTT	3528
O	y	598	GCAAAAGGGTATGATTAAGTCAGCACTTAATAGATTAGTATGATGGGATAGCTGGCCA	657
D	b	3525	GCAAAAAG --- CCGGACGTTTGGGACAGCTGGCCA	3493
O	y	658	CTGCAAAAACGCTGATGACTGCGGCAAAATATCATGGATATPCATCGTCTCCGCTTTA	717
D	b	3492	TTACAAAAACGCTGACGGCAGCTGTGGCAAACTATCACGGCTACCAATCGTCTTTGCAATTA	3433
O	y	718	GCAGGTGACCCAAAACACAGTGAATGATCTCACTCATTTATTCTATCAAAAAGTCGT	777
D	b	3432	GCCGAGATCTTAAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGTCGGC	3373
O	y	778	GATCATCGATTTTGACAGCTGGAAAAATGCTGGAAAGATTTGGAAGATGGAATGATTT	837
D	b	3372	GAACCTCTATTTGACAGCTGGAAAAACGCTGCGCGCTTTAAAGACAGCGACAAAATTC	3313
O	y	838	GTTCCAATGATGCCGATCTTAATATATCAACACAGAGTGGTCAAGCTTCGTACTTTA	897
D	b	3312	GATGCAAAATGATWTCTATCTCTAAAGACCAACACAGATGCTGAGGTTCCAGCACATTT	3253
O	y	898	ACCAAGATGGCCAGTCCGTTTCTCATACATTAACATACAGGTAATCCGAAGATGAT	957
D	b	3252	ACATCTGACGAAAAATCCGTTTATTTCTACCTGATTTCTCCGGTAAACATTA ---	3200
O	y	958	GGAACCGGTGCTGGTAAACCAATCATTTTCACTGCTCAAGTAACCTATCCACCGGAT	1017
D	b	3199	-----CGGCAAAACAAACACTGACCACTGCACAGTTAAGTATCAGCATCAGAC	3151
O	y	1018	GCACCTACACTTAAAGTCAGTGAGTATCTGATATTAATCTGCTTTGATGGCGGAGAC	1077
D	b	3150	AGCTCT --- TTGAACATCTAACGGGTGTAGAGGATTTAATAATCAATCTTGA ---	3097
O	y	1078	GGTACAGTTTATCAAAATATTCACACAATTTATTCATGAAGGCAAGTGGATTTACAGTAT	1137
D	b	3096	GGAANAAGTATCAAAATGATACACAGTTCATGATGAAGCACTACAGCTCAGGGCAC	3037
O	y	1138	AACCATCTTTAAGAGACCTCACTATGTTGAAATGAAGGCCATAATATCTTGTCTTT	1197
D	b	3036	AACCATCTCGTGAAGATCTCTCACTACGTAAAGATTAAGGCCCAAAATACTTGTATTT	2977
O	y	1198	GAACGGAATCTGGAACAACAGATGATTATACAGGCATCAGTCTTTCATAAATAAAGCT	1257
D	b	2976	GAACGAACACTGGAACATGAAGATGGCTACCAAGCGCAAGATCTTTATTTAAACAAAGCA	2917
O	y	1258	TACTATGGCGGAGTGAAGCTCTTCTCCAGATGAAAAAAATAATGATGCTTCAAGTCTT	1317
D	b	2916	TACTATATGGCAAAAACACATCACTTCTCCGCTCAAGAAAGTCAAAAACCTTCCCAAGCAT	2857
O	y	1318	AAAAAACAAATATGCTCTTATAGCAATGATGTCATTAAGCATTTGTAATTGGCGGATGAC	1377
D	b	2856	AAAAAAGCACGCGCTGAGTTAGCAAAACGGCGCTCTCGATGATTTAGACTAAACGATAT	2797
O	y	1378	TATACAGTGAAGAACTGTTATGAACCAATATAGTCACAACACAGTACAGCAAGTCAATC	1437
D	b	2796	TACACACTGAAAAAAGATGATGAACCCGCTGATTCATCTAACAACAGTAAACGATGAAT	2737
O	y	1438	GAACGGCGCAATATATTTAAAAATGAATTAATATGATATCTATTACAGGATTTCAAGAGA	1497
D	b	2736	GAACGGCGCAACGCTTTTAAATTAACAGCGCAAAATGATATCTGTTCACTGATCCCGCGGA	2677



OY	1498	TCCAAATAGCAGTSGATGGAATTATACGACAAGAATGTTATATATGCTAAGGCCCGGAGGC	1557
Dd	2676	TCAAAATAGCAGTATGACGGCATTACGCTTAGACATAATTTACATGCTTGCTATGTTC	2617
OY	1558	GACTCCTTAATGGCCCCACACACCAGCATTAATGAACCTGCACTTGATATGACATGAAT	1617
Dd	2616	AATTCCTTAATGGCCCCATACAAACCGGTGACAAAACTGGCCCTTGCTTTAAATATGAT	2557
OY	1618	CTTAGACCTGCTGATTCACACACACTTACTCTCATTTGGCGGTATCCGCAACCTGAANGT	1677
Dd	2556	CTTATCTTAACGATGTAACCTTTACTACTCACACTTCGCTGACCTCAAGCCAAAAGA	2497
OY	1678	AATATGATGGATCACACAGTATATATGACGAATPAGGCTTCPTACGACATCACTCT	1737
Dd	2496	AACATATGCTGATATTACAGCTATATATGACAAACAGAGATTTCTACGACAAACAATCA	2437
OY	1738	CACCTGGCGGACAAGCTTGGGTTAATATTAAGGCTGACACATCTGAGAGGAAAT	1797
Dd	2436	ACGTTTGGCGCTACTCTTCCTGCTGAACATCAAAGCAAAAGAACATCTGCTGTCAAGAC	2377
OY	1798	AGTTCGGACACAGACAAATTCCTCCATAGCGATTATCTCCCATAAAGAATATGCACTGG	1857
Dd	2376	AG---CATCTCTGAACAAGACAAATTAACACTTAACAAATTAACCAACCAAGAAAAATGC	2320
OY	1858	CAGATGCTCACTTTTCCCAACCTTTTTTTTTTAAATAAGGCTTTGTACACACTTA	1917
Dd	2319	CGATATCTCTAATGGCAATTTCTTTATTTATTTCTTATCAACATTAAGTGTAATCCCATATGAA	2260
OY	1918	GAGAAAAAAGAA	1930
Dd	2259	CTATATAAAGCA	2247

RESULT 15				
BS294043/c				
LOCUS	BS294043	88063 bp	DNA	linear
DEFINITION	B.subclillis genomic DNA fragment (88 kb).			
ACCESSION	Z94043			
VERSION	Z94043.1			
KEYWORDS	GI:1945641			
	amino acid racemase; CRH gene; levansucrase; para-nitrobenzyl			

```

SOURCE      Bacillus subtilis.
ORGANISM    Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE   1 (bases 1 to 88063)
AUTHORS     Denizot,F.C.
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 88063)
AUTHORS     Denizot,F.C.
TITLE        Direct Submission
JOURNAL      Submitted (11-APR-1997) F.C. Denizot, C.N.R.S., Laboratoire De
              Chimie Bacterienne, 31, Chemin J. Aiguier, Marseille Cedex 20,
              13402, FRANCE
FEATURES
  source      location/Qualifiers
              1..88063
              /organism="Bacillus subtilis"
              /db_xref="taxon:1423"
              1..196
gene

```

```

/gene="yvca"
<1..196
CDS
/gene="yvca"
/note="Similar to SUGC_ECOLI sugc protein"
/codon_start=2
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB08070.1"
/db_xref="GI:1945642"
/db_xref="SEPTREMBL:O06965"
/transl_table="KYNYSYSYGVKSTEDNIGVEAYFTLGCKNENQPCSIDDEEKA
KKVVEKLINSITFLDKRKK"
198..2108
gene
/gene="yvcB"
198..2108
CDS

```

```

/codon_start=1
/translate_table=11
/product="hypotheical protein"
/protein_id="CA808071.1"
/db_xref="GI:1945643"
/db_xref="SETRMBL:006966"
/translation="MGNKNVFNENNVKLRLLIDLEYDYKQKFASTIASEKKAKKDFEAD
DIRRIYKLETHHDLPENNKLEIYISNELLVONTSIOSSTRSGYDGAIIHIDEKKHII
DOLQIISGSANDKMSYNNPGLFLGIDHSOYEATKEFTNNAKAANSBULKTEFPALG
SLANNNOVMOLIDEGEDVYGVNGVAOVNIDOLETDESILYRVIRKAFSVKKEDDS
IPELKEKKEIKYVEDKGVNTINJROTSKDDPLIGVSGSKAPFTFFGVKKNADNTSVK
GIRINMDGVPPEDNRSIOKTELOKTIKODIYEKQGLGFLKAASGIDIDILEKKNNTDGL
KKVSLVYLFHFDQKQMGRIKEELPELPSGFLMLLRNSGPAVDVDAVNGSIIDEOKKVV
IKKELITINASNKRGIESRYEHRIDHLLGQFQOMALKDGEIVLADKSTLSLKLTDLE
TDDAKLITSHISIVOMLNAISKEGFSFYKCSGDIYFNGSGSGETIOVNLSASARIYO
NGMKLVEMEDDISKYOKVYOSIEDIDEDIDKKOALITAIIHMEENPSHVAFNDOLFRIA
ACGSTRPKLEKISVHESFHTGALPANDGIYVLELKQYTEKNDITKNIRESIEKLFKE
EEMISQLDFEOS"
2647..4416
/gene="yvcC"
2647..4416
DSD

```

/note="probable ABC transporter"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAB08051.1"  
 /db\_xref="gi:1945644"  
 /db\_xref="SPTREMBL:006967"  
 /translation="MPTKKOKSKSKIKFFELVARTNPSTYKLAFLALSLVYTVLSEL  
 LIPILTKLOVDFSMNSLSTGIGLIALVFFVQGLSVATYALNTNGKISLRELL  
 LMKRLIKLPVSPFDINASGETVSRVNTDTPVYKLLITHTLSGFTIGLIIVSIGSTLTILE  
 LNMNMLTILVAVVLEALILVPIGRKMFSSIRETQDETARFTILNQILPEIRLVVAA  
 SNAEDEVGGRGKMGISLSFLKLVREAKVQSLVGLISLVLMALVAVNIAGYGMQVSSG  
 ELTVALVAFILLYLEFQITLPMGOITPTFLOKSGTATERNIELIAEEDDEPTVYKQI  
 ENAHLPIDLDREYFGYKPDOLIKREVSVAIVAGKVTAVVGGSGGKTTLPKLPERS  
 PAAGTIRIGDEPDVPSLESREHNTGYVSSQSPSLMSGTIRNITCYGLERYVTDAIEIK  
 AAEBYALNFIKELPNQDTEVGERHGLMSGQKRIATIALLRNPSITLMDLPAVSS  
 LDSQSKSVQOALEVLEMGRTTIVIAHRLSTVVDADOLLFEVKEGELTGRGTHHELMAS  
 HGLYDFAEQQLKMNADLENKAG"  
 4542..5996  
 /gene="yvcd"  
 4542..5996  
 /gene="yvcd"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAB08052.1"  
 /db\_xref="gi:1945645"  
 /db\_xref="SPTREMBL:006968"  
 /translation="MGKHTSEHKHNAQVYVOLLDDGYEFPHKGLKAYKERNIKRASKLI  
 ORAVHLPEPDESEMLISOLAVIYSEMGQOVSQEDLIDYIMANIEMPECHCYFKANPNF  
 LGLPQETAKKAAYSADADPDGEFAEENDSLDLDDGDESLVDODELTKODRA  
 KSLDESGGLAEVVALEETITTEYPEDLSAVNNLALAYFGSNVAKAKQIVALEVLSKE  
 GNULHMLCLVYYIYERDEKVAELISQLSNMYMLLEQRKILGATLALVAGIEIGK  
 WLKSYLKGFGEDDPEFFWYMLSCSAVFETGHDFAETIRKHYESQVPGEDRPAAYEIK  
 ALPSSVEORFLAAVYIISRTKETHELVAVIRSKRTIAPETKVFYVITLNGSQAADVSE

```
gene
CDS
    /protein_id="CAB08053.1"
    /db_xref="GI:1945646"
    /note="similar to p54entfc p54 protein precursor"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAB08053.1"
    /db_xref="SPTREMBL:O06969"
    /translation="MRSLITGLASVIGTSFLIPFTSKTASAEFLDEKKOKIESKO
    SEVASIEAKEKELELOENOSKIEKLDINDKALDINSKTEDEKKEENDKKEIKK
    LKKEIKETAREKRENETIKRKVRSLQSGSGOYIDVLGTSFGDFISRTAVYSI
    VDADKDLKQODKAKLEDSADLNDKREYQALAKLETQKODLQDNKEDLFD
    EAKASQKTAKAISELSEASLANOKANTEAOARKKEDEAAALIKOEAQAKS
    DETQNDSDOTATESSKASSSDSDNSDNSGNSGSSSSSKSSKSSGNSGNSG
    TVISNSGIEGAIYVSSIVGOSPYKFGGFRQSDINNRIFCDSFVRVAYASAVNL
    GPVGCTTDTLVRGQAVASASEMRGDLVFPTTYKNGHVIGLNGFELNDNTSHV
    SVDMSNPYKRAFKGVVRVVO"
    complement(7228..7380)
    /gene="yvcf"
    /complement(7228..7380)
    /gene="yvcf"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAB08054.1"
    /db_xref="GI:1945647"
    /db_xref="SPTREMBL:O06970"
    /translation="MPPELLFPELLFLEDEPDELDEPDELSELSSESDLEAFD
    DSVAVV"
    8004..8954
    /gene="yvcH"
    8004..8954
    /gene="yvcH"
    /note="probable thioredoxin reductase"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAB08055.1"
    /db_xref="GI:1945648"
    /db_xref="SWISS-PROT:P80880"
    /translation="MSEKIDYVILIIIGAGPAGTAAYTSRANLSTLMIERGIPGGM
    ANTEDEVNPGFESILPELSNKMFEHKKFGAAYAGDIKVIDGKEKVKYAKASKE
    KYARAVIIAAGAEYKIKVPGKEKEIGKRVSYCAVDGAFKGLVYVGGDSAAVE
    GYTLRFASKVTIVHRDKLRQSTLQARAFDNEKVDLFNKTVEIHEENKVENVT
    LVDIVTGESEKRTGDFVFIYIGMLPLSKPFEMLGITNEGYIETNDRMETKVEGIFAA
    GDIRKSLRQIYATATGDSIAQSVQHYVEELQETLTKL"
    9272..9748
    /gene="yvcI"
    9272..9748
    /gene="yvcI"
    /note="probable nudix hydrolase"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAB08056.1"
    /db_xref="GI:1945649"
    /db_xref="SPTREMBL:O06972"
    /translation="MTYLQRYNVCVLQETDDKVLLOKPRGMVAPGGMESGESVRD
    SVIREYRETGIIYINPOLKGVFTFLIKGDHIVSEMMFTFVADSYGONVSESEEG
    KLOHMDVNDIONLPMAPDGHILDFPMKMGQGLLHGTFYTPPEFELLSTRLDQHK"
    9773..10660
    /gene="yvcJ"
    9773..10660
    /gene="yvcJ"
    /note="similar to hypothetical MTCY21B4"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
```

```
Query Match      25.4%; Score 611.8; DB 1; Length 88063;
Best local Similarity 65.1%; Pred. No. 1.7e-128;
Matches 1024; Conservative 0; Mismatches 492; Indels 57; Gaps 6;

gene
CDS
    /protein_id="CAB08057.1"
    /db_xref="GI:1945650"
    /db_xref="SWISS-PROT:O06973"
    /translation="MSVSESHDIQVITITGSGAGKTYAATOSFEDLGFVCDNLPRL
    LPELELMKESNSKSWKVALVMDLVEKREFEFLLEADEMENPITRIILFLDAKDS
    ILVRYKETEKRSHPLAATGLPEGLALEKRSQIITVDSMKPRDLREKIV
    KHFAVNOGETFVYVNVMSFGFYGIPIIDDLVDFVDFPNPYIISMRLTCKDEVS
    YVMKNVETOKENERKLIDLSFMLPSPYKREGKSQVYVIAIGTGGHRSVTLAENLADYF
    KDDYTHVTHRDIEKRSRK"
    10662..11615
    /gene="yvcK"
    10662..11615
    /gene="yvcK"
    /note="similar to hypothetical SYCSLRA"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"

338  ACGATGAACCTCAAAAAGATTGGCGAAAAAGCAGCTGCCGTAACCTTGACGACTCTATA 417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
46152 ACGATGAACATCAAAAAGCTTTGCAAAACAGCAACAGATTAACCTTACTACCGCACTG 46093
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
418  TTAGTAGAGGCGGAGCGACCGCATATTTTGGCGCAAAATGCAACAGCGGGGACTACAG 477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
46092 CTGGCAGAGG--CGAAGCTCAAGCGTTTGGAGAAGAACCAACCAAGCATATYAG 46036
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
478  GAAGACTATGCTTTTGGCCATATACAGCGCGTGACATGCTAAAATTCACAGAACAA 537
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
46035 GAAACATNCGGATTTCCCATTTACAGCCATGATATGCTGCAAAATCCGTAACAGCAA 45976
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
538  AACAGTCTCAATTAAGTGCCTCAATTCATGATGATGATGATGATGATGATGATGATG 597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45975 AAAAATGAAAATATCAAGTTCTTCGATTCGATTCGATTCGATTCGATTCGATTCG 45916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
598  GCAAAAGGTATGATAGTACAGCAACTTAATAGATTAGATGATGATGATGATGATGATG 657
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45915 GCAAAAG-----CTGGAGCTTTGGGACAGCTGGCCA 45883
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
658  CTGCAAAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 717
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45882 TTACAAAACGCTGAGCGGACCTGTCGCAAACTATACAGCGGTACACATCGTCTTGCATTA 45823
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
718  CGAGGTGACCCAAAAACAGATATATATCTCATTCTATTTCTATCAAAAAGTGGT 777
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45822 GCCGAGATCCTAAAAAATGCGGATGACATCGATTTCATGTTCTATCAAAAAGTGGC 45763
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
778  GATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 837
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45762 GAACTTCTATTTGACATGCTGAAAAACGCTGGCGCGCTTTAAAGACAGCACAATTC 45703
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
838  GTTCAAAATGATCGTATCTTAATATCAAAACAGAGAGTGTGAGTTCTGCTACTTTA 897
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45702 GATGCAAAATGATTTATCTTAAGACCAAAACAGAAATGCTGAGTTTACACCCACATTT 45643
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
898  ACCAAAGATGCGCAGTCCGTTTATCTATACAGATTACTGAGTATTCGAAATGCT 957
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45642 ACATCTGACGGGAAAATCGGTTTATCTACATGATGATTCGCGTAACATTA----- 45590
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
958  GGAACCGGTGCTGTAACCAAAATCAATTCAGTGTCAAGTAATTCACAGCGGAT 1017
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45589 -----CGGCAAAACAACTGACACTGCAACGTAATACGTATACAGATCAGAC 45541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1018  GCAGCTACACTTAAAGTCAGTACGATATCTGATCAATTAATCTGTTTATGATGCGGAGAC 1077
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45540 ACCTCT---TTGACATCATCAGCGTGTAGAGATTATTAATCAATCTTTGA---CGGTGAC 45487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1078  GGTACAGTTTATCAAAATTTTACGATTTATCGATGAAGGACAGTGGATTTACAGTAT 1137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45486 GGAATAAGCTATCAAAATGATACAGCATGATGATGATGATGATGATGATGATGATGATG 45427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
OY 1138 AACCATCTTAAAGACCTCTACTATGTTGAAGATAAGGGCATAAATATCTTGTCTT 1197
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45426 AACCATACGCTGAGAGATCCTCACTACGTAGAGATGAAGGCCACAATACTAGTATT 45367
OY 1198 GAAGCGAATCTGGAACAGATGTTATCAAGCGATCGTCTTCAATATAAAGCT 1257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45366 GAAGCAACACTGGAAGTGAAGTGGCTACCAAGCGAAGATCTTATTTTACAAAGCA 45307
OY 1258 TACTATGCGGAAGTGAAGTCTTCTCTCAGATGAAAAAATAAAGTCTCAAGTCTT 1317
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45306 TACTATGCAAAACACATCATTTCTCCGTCAGAAAAGTCAAAAACCTTCTCAAAAGCAT 45247
OY 1318 AAAAAACAAATTCCTTCTTAAAGCAATGTCATTTAGCATTTGTAATTTGCCGATGAC 1377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45246 AAAAAACGACGCTGAGTTAGCAAAAGCGCCCTCGGTATGATTTAGCTAAAGCATGAT 45187
OY 1378 TATACATGAAAAAGTGTATGAACCATAGTCCATCAAAACAGTAGACAGATGAAGTC 1437
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45186 TACACACTGAAAAAGATGATGAAGACGCTGATTTGATCTTAACACAGTAAACGATGAAAT 45127
OY 1438 GAACGCCCAATATATTTAAATGAATAAATGATATCTATTCACGATTCAGAGGA 1497
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45126 GAACGCCGGAACGCTTTAAATGAACGCAATGATGTAACCTGTTCACTGACTCCCGCGGA 45067
OY 1498 TCCAAATGACGAGTGAATTAACGACAAGATGTTTATGCTAGAGCCCGGAGGC 1557
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45066 TCAAAATGACGATTTAGCGGATTAAGCTTAACGATATTTCATGCTGTATGTTTCT 45007
OY 1558 GACCTCTTAATGGCCCAACACACACCATTAATGAACGTGACTTGTATGACATGAAAT 1617
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45006 AATCTTTAACTGGCCATACAGCGCTGAACAAACCTGCGCTGTGTAAAAATGAT 44947
OY 1618 CTTGACCTCTGATCTCACACACACTTACTCTCATTTGGGTATCCGACCTGAAGGT 1677
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44946 CTTGATCTTAAGCATGTATACCTTACTTACTACACACTTCGCTGTACTCAAGCGAAGGA 44887
OY 1678 AATAATGTGTTACTCACAAGTTATATGAGATAGAGGCTTCTATCCAGAAATCACTCT 1737
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44886 AACAAATGCTGATTAAGATATGACAAACAGAGATTTCTACGACACAAACAAATCA 44827
OY 1738 CACCTGGGGGACAGCTTGGGGTTAATATTAAAGGCTTGACACATCTGAGGAGAAAAT 1797
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44826 ACGTTTGGCCCAACCTTCTGCTGAGACATCAAAAGCAAGAAACATCTGTGTCAAAAGAC 44767
OY 1798 AGTTCCGACAAGACAAATTCCTCATAGCATTTATCTCCAAATAAAGAAATGTCACTGG 1857
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44766 AG---CATCTTGAACAAGACATTAACAGTTAACAATTAATAAGCAAAAGAAATGC 44710
OY 1858 CAAGATCCTACCTTTTCCCAACCTTTTAAAAATAAAGGTTTGTACCACTTTA 1917
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44709 CGATATCTCTATTGCAATTTCTTTATTTCTTATCAACATTAAGGTGAATCCCATATGAA 44650
OY 1918 GAAGAAAAAAGAA 1930
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44649 CTATATTAAGCA 44637
```

Search completed: June 18, 2003, 07:25:36  
Job time : 6282.52 secs

